

UNIVERSITÉ DU QUÉBEC

MÉMOIRE PRÉSENTÉ À
L'UNIVERSITÉ DU QUÉBEC À TROIS-RIVIÈRES

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN MATHÉMATIQUES ET INFORMATIQUE
APPLIQUÉES

PAR
CLÉMENTINE VASSOILLES

PROPOSITION D'UNE NOUVELLE MÉTHODE DE CLASSIFICATION
À BASE DE COPULES

JANVIER 2014

Université du Québec à Trois-Rivières

Service de la bibliothèque

Avertissement

L'auteur de ce mémoire ou de cette thèse a autorisé l'Université du Québec à Trois-Rivières à diffuser, à des fins non lucratives, une copie de son mémoire ou de sa thèse.

Cette diffusion n'entraîne pas une renonciation de la part de l'auteur à ses droits de propriété intellectuelle, incluant le droit d'auteur, sur ce mémoire ou cette thèse. Notamment, la reproduction ou la publication de la totalité ou d'une partie importante de ce mémoire ou de cette thèse requiert son autorisation.

REMERCIEMENTS

Je souhaite remercier l'ensemble des personnes qui ont participé, de quelque manière que ce soit, à ce mémoire. Tout d'abord, j'aimerais exprimer toute ma reconnaissance envers mes deux directeurs de recherche, MM. Mhamed Mesfioui et Jean-François Quessy, tous deux professeurs au département de Mathématiques et d'Informatique de l'Université du Québec à Trois-Rivières. Sans eux, ce travail n'aurait pu voir le jour. Je les remercie de m'avoir épaulé, dirigé et conseillé tout au long de ces deux années à la maîtrise. J'ai beaucoup apprécié travailler avec eux.

J'aimerais également remercier ma famille. Tout d'abord, mes parents, Frédérique et Jean-Pierre, ainsi que mon frère Pierre, qui ont toujours su trouver les mots pour m'encourager et me réconforter malgré la distance qui nous sépare. Je remercie aussi mon oncle et ma tante, André et Claude, sans qui cette expérience québécoise n'aurait jamais pu avoir lieu ; merci d'avoir confiance en moi. Je tiens aussi à exprimer toute ma gratitude à mon amoureux, Michaël, pour ses encouragements, sa bonne humeur ainsi que son soutien inconditionnel, même dans les moments difficiles.

Mes études à la maîtrise ont été financées en partie par des subventions de recherche octroyées à MM. Mhamed Mesfioui et Jean-François Quessy par le *Conseil National de Recherche en Sciences Naturelles et en Génie* du Canada. Je remercie également l'*Institut des Sciences Mathématiques* du Québec, ainsi que l'Université du Québec à Trois-Rivières pour les bourses d'étude qu'ils m'ont accordées.

Table des matières

Remerciements	ii
Liste des tableaux	vii
Liste des figures	viii
Chapitre 1. Introduction	1
Chapitre 2. Probabilités et statistiques	4
2.1 Notions de probabilités	4
2.1.1 Définition d'une probabilité	4
2.1.2 Probabilité conditionnelle	5
2.1.3 Formule de Bayes	6
2.2 Les variables aléatoires	6
2.2.1 Variables aléatoires discrètes	7
2.2.2 Variables aléatoires continues	9
2.3 Lois de probabilités discrètes	11
2.3.1 Distribution binomiale	11
2.3.2 Distribution de Bernoulli	12
2.3.3 Distribution géométrique	12
2.3.4 Distribution hypergéométrique	13
2.3.5 Distribution de Poisson	13

2.4	Lois de probabilités continues	14
2.4.1	Distribution normale	14
2.4.2	Distribution Gamma	15
2.4.3	Distribution de Weibull	15
2.4.4	Distribution Beta	16
2.4.5	Distribution lognormale	16
2.5	Les vecteurs aléatoires	16
2.5.1	Les vecteurs aléatoires discrets	16
2.5.2	Les vecteurs aléatoires continus	18
2.5.3	Covariance et corrélation	19
Chapitre 3. Méthodes de classification bayésienne		21
3.1	Introduction	21
3.2	Théorie	22
3.3	Loi Normale sous l'indépendance	24
3.4	Loi Normale avec matrices de variance-covariance homogènes .	25
3.5	Loi Normale avec matrices de variance-covariance hétérogènes	27
Chapitre 4. Copules : théorie et estimation		30
4.1	Définition et Théorème de Sklar	30
4.2	Propriétés des copules	33
4.2.1	Invariance d'une copule	33
4.2.2	Copule d'indépendance et bornes de Fréchet	34
4.3	Exemples de copules	35
4.3.1	Copule Normale	35
4.3.2	Copule de Student	36
4.3.3	Copules Archimédiennes	37

4.4	Mesures de dépendance	40
4.4.1	Tau de Kendall	41
4.4.2	Rho de Spearman	43
Chapitre 5. Une méthode de classification à base de copules		44
5.1	Objectif	44
5.2	Méthodologie	45
5.3	Modèle à base de copules Archimédiennes	46
5.3.1	Description générale	46
5.3.2	Exemple avec la copule de Clayton	46
5.3.3	Effet des marges sur l'affectation	48
5.4	Probabilités d'erreur de classification	49
5.4.1	Cas à deux classes	50
5.4.2	Cas à K classes	53
5.5	Illustration sur de vraies données	54
Chapitre 6. Conclusion		57
A. Programmes de simulation en Matlab		59
A.1	Estimation des probabilités d'erreurs	59
A.2	Estimation de paramètres par inversion du tau de Kendall . . .	60
A.3	Simulation de données bivariées selon une certaine copule . . .	64
A.4	Transformation d'un échantillon univarié en rangs	69
A.5	Calcul des fonctions discriminantes	70
A.6	Classification d'une nouvelle observation	70
A.7	Calcul de la fonction discriminante	71
A.8	Algorithme de classification d'un jeu de données	72
A.9	Calcul du taux d'erreur de classification	73

TABLE DES MATIÈRES	vi
--------------------	----

Bibliographie	59
----------------------	-----------

LISTE DES TABLEAUX

5.1	Estimation des probabilités d'erreurs de classification pour la copule de Clayton	53
-----	--	----

LISTE DES FIGURES

3.1	Exemple de règle de décision bayésienne	29
4.1	En haut : densités de la copule Normale quand $\rho = .5$ (à gauche) et $\rho = .9$ (à droite); en bas : courbes de niveau de la copule Normale quand $\rho = .5$ (à gauche) et $\rho = .9$ (à droite)	36
4.2	Densité (à gauche) et courbes de niveau (à droite) de la copule de Student à 1 degré de liberté et $\rho = .5$	37
4.3	Densités de la copule de Clayton pour $\theta = 1$ (à gauche) et $\theta = 3$ (à droite)	40
4.4	Densités de la copule de Frank pour $\theta = 1$ (à gauche) et $\theta = 3$ (à droite)	41
5.1	Seuil de décision pour deux copules de Clayton avec $\theta_1 = 2, 9$ et $\theta_2 = 2, 5$	48
5.2	Seuil de décision pour deux copules de Clayton avec marges exponentielles	50
5.3	Nuage de points des rangs standardisés du groupe Ω_1 (à gauche) et Ω_2 (à droite)	55

CHAPITRE 1

INTRODUCTION

Soit une population d -dimensionnelle Ω partitionnée en K classes, que l'on note $\Omega_1, \dots, \Omega_K$. Autrement dit, $\Omega = \Omega_1 \cup \dots \cup \Omega_K$ et $\Omega_k \cap \Omega_{k'} = \emptyset$ pour tous $k \neq k' \in \{1, \dots, K\}$. On suppose que ces classes sont connues *a priori*. Supposons que l'on observe la réalisation d'une variable aléatoire $\mathbf{X} \in \mathbb{R}^d$ et que l'on souhaite déterminer dans quelle classe parmi $\Omega_1, \dots, \Omega_K$ elle appartient. Cette procédure s'appelle la *classification*.

Une des méthodes de classification les plus utilisées est celle basée sur des fonctions discriminantes. Elle a été décrite abondamment dans la littérature. L'idée consiste d'abord à évaluer la probabilité que \mathbf{X} appartienne à chacune des classes $\Omega_1, \dots, \Omega_K$. En d'autres termes, on considère pour chaque $k \in \{1, \dots, K\}$ la fonction $g_k(\mathbf{X}) = P(\Omega_k | \mathbf{X})$. Par une application de la formule de Bayes, on peut écrire

$$g_k(\mathbf{X}) = \frac{P(\mathbf{X} | \Omega_k) P(\Omega_k)}{\sum_{k=1}^K P(\mathbf{X} | \Omega_k) P(\Omega_k)}, \quad (1.1)$$

où $P(\mathbf{X} | \Omega_k)$ est la probabilité d'observer \mathbf{X} sachant que la loi de \mathbf{X} est représentée par la distribution dans la classe Ω_k ; cette probabilité s'appelle

la *vraisemblance*. L'observation \mathbf{X} sera alors affectée à la classe dont la vraisemblance est la plus élevée. Puisque le dénominateur à l'équation (1.1) est le même pour chaque fonction discriminante, il s'agit de choisir la classe k telle que $P(\mathbf{X}|\Omega_k)P(\Omega_k)$ est la plus élevée parmi

$$P(\mathbf{X}|\Omega_1)P(\Omega_1), \dots, P(\mathbf{X}|\Omega_K)P(\Omega_K).$$

Dans la littérature classique, les fonctions discriminantes sont habituellement basées sur la loi Normale multivariée, notée $N_d(\mu, \Sigma)$, où $\mu \in \mathbb{R}^d$ est le vecteur moyenne et $\Sigma \in \mathbb{R}^{d \times d}$ est la matrice de variances-covariances. Sa densité de probabilité est donnée par

$$h(\mathbf{x}) = \frac{|\Sigma|^{-1/2}}{(2\pi)^{d/2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mu)^\top \Sigma^{-1} (\mathbf{x} - \mu) \right\}, \quad \mathbf{x} \in \mathbb{R}^d.$$

Les fonctions discriminantes dans ce contexte sont alors

$$g_k(\mathbf{X}) = -\frac{1}{2} (\mathbf{X} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{X} - \mu_k) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\Sigma_k| + \ln P(\Omega_k),$$

où μ_k et Σ_k sont respectivement la moyenne et la matrice de variances-covariances de la classe k . En supposant que les probabilités $P(\Omega_1), \dots, P(\Omega_K)$ sont identiques et en faisant abstraction du terme constant, on peut prendre en considération les versions modifiées de g_1, \dots, g_K , à savoir

$$\tilde{g}_k(\mathbf{X}) = -\frac{1}{2} (\mathbf{X} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{X} - \mu_k) - \frac{1}{2} \ln |\Sigma_k|, \quad k \in \{1, \dots, K\}.$$

Une faiblesse majeure de cette approche est que l'on doit supposer que les observations sont distribuées selon le modèle de la loi Normale, ce qui est loin d'être toujours le cas en pratique. Ainsi, s'il s'avère que l'hypothèse de normalité des observations à classer n'est pas respectée, on peut se retrouver avec de graves erreurs de classification.

L'objectif principal de ce mémoire est de pallier à cette faiblesse en proposant une approche plus générale basée sur les copules. Ainsi, au lieu de supposer une loi Normale d -dimensionnelle, nous allons plutôt baser notre approche sur des lois générales construites à base de copules. Un cas particulier de notre approche sera la méthode décrite dans le paragraphe précédent. Notre méthode est originale puisqu'elle n'a jamais été considérée jusqu'ici. De plus, les méthodologies développées dans ce mémoire sont simples du point de vue de leur implémentation et demandent un faible temps de calcul.

Le reste du mémoire est organisé comme suit. Le Chapitre 2 fait un survol des méthodes de classification probabilistes. Le Chapitre 3 concerne les différentes lois de distributions des individus, alors que le Chapitre 4 recense quelques-uns des résultats les plus importants concernant la théorie des copules. Le Chapitre 5 constitue l'apport original de ce travail de recherche. On y introduit d'abord formellement notre nouvelle méthode de classification basée sur les copules. Ensuite, on montre et illustre comment estimer la probabilité d'erreur de classification et on présente en détail un exemple d'applications sur des données relatives au contrôle informatisé de billets de banque. Le mémoire se termine par une brève conclusion. Tous les programmes Matlab qui ont permis de mettre en œuvre notre méthode sont présentés à l'Annexe A.

CHAPITRE 2

PROBABILITÉS ET STATISTIQUES

2.1 Notions de probabilités

L'ensemble des sections présentées dans ce chapitre sont issues du livre *Probabilités, statistique et applications* écrit par ?.

2.1.1 Définition d'une probabilité

Intuitivement, on définira une probabilité comme étant la chance qu'un événement se produise. En voici une définition plus formelle énoncée par ?.

Définition 2.1. Soit (Ω, \mathcal{A}) , un espace mesurable, où Ω est l'ensemble fondamental et \mathcal{A} , une partition de cet ensemble. On appelle probabilité sur (Ω, \mathcal{A}) toute mesure positive $P : \mathcal{A} \mapsto [0, 1]$ telle que $P(\Omega) = 1$. De plus, pour toutes suites d'événements mutuellement disjoints A_1, A_2, \dots dans \mathcal{A} , on a

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

Voici maintenant quelques propriétés d'une mesure de probabilité.

Propriété 2.1.

- (i) La probabilité de l'événement impossible est 0 ;
- (ii) La probabilité de l'événement complémentaire de A , à savoir $\bar{A} = \Omega \setminus A$, est $P(\bar{A}) = 1 - P(A)$;
- (iii) La probabilité de l'union de deux événements A et B est donnée par

$$P(A \cup B) = P(A) + P(B) - P(A \cap B);$$

- (iv) Si deux événements A et B sont incompatibles, i.e que $A \cap B = \emptyset$, alors

$$P(A \cup B) = P(A) + P(B).$$

2.1.2 Probabilité conditionnelle

Définition 2.2. Soient A et B , deux événements dans \mathcal{A} tels que A est un événement de probabilité non nulle. On appelle probabilité conditionnelle de B par rapport à A la probabilité de la réalisation de B sachant que A s'est produit. On note cette probabilité

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$

On déduit de la définition précédente que

$$P(A \cap B) = P(A) P(B|A) = P(B) P(A|B).$$

2.1.3 Formule de Bayes

Certains événements peuvent être décomposés en deux parties. En effet, il est possible de calculer les probabilités *a posteriori* $P(A_i|B)$ après suite à la réalisation d'un événement B à partir des probabilités à priori $P(A_i)$. Cette probabilité est calculée à l'aide de la formule de Bayes.

On suppose d'abord une partition de Ω en événements A_1, A_2, \dots, A_n . Ainsi, $A_i \cap A_j = \emptyset$ pour tout $i \neq j \in \{1, \dots, n\}$ et $\Omega = \cup_{i=1}^n A_i$. On suppose aussi que $P(A_i) > 0$ pour tout $i \in \{1, \dots, n\}$. À partir d'un certain événement B , on suppose connues les probabilités conditionnelles $P(B|A_i)$, $i \in \{1, \dots, n\}$.

On note d'abord que

$$B = B \cap \left(\bigcup_{i=1}^n A_i \right) = \bigcup_{i=1}^n (A_i \cap B).$$

Puisque les événements $A_1 \cap B, \dots, A_n \cap B$ sont disjoints, on déduit

$$P(B) = \sum_{i=1}^n P(A_i \cap B) = \sum_{i=1}^n P(A_i) P(B|A_i).$$

De là, on obtient la formule de Bayes, à savoir

$$P(A_i|B) = \frac{P(A_i \cap B)}{P(B)} = \frac{P(A_i) P(B|A_i)}{\sum_{i=1}^n P(A_i) P(B|A_i)}.$$

2.2 Les variables aléatoires

Une variable aléatoire est une fonction à valeurs numériques définie sur un espace échantillon. En d'autres termes, c'est une fonction définie sur l'en-

semble des résultats possibles d'une expérience aléatoire. Elle peut être de nature discrète ou continue.

2.2.1 Variables aléatoires discrètes

Une variable aléatoire est dite discrète si le nombre de valeurs qu'elle peut prendre est fini ou infini dénombrable.

Définition 2.3. *On appelle variable aléatoire discrète sur (Ω, \mathcal{A}) une fonction $X : \Omega \mapsto \mathbb{R}$ telle que $X(\Omega)$ est dénombrable.*

La loi de probabilité d'une variable aléatoire est représentée par la probabilité associée à chacune des valeurs possibles de X . Ainsi, la probabilité que la variable aléatoire X prenne la valeur x est donnée par

$$P_X(x) = P(X^{-1}(x)) = P(\omega \in \Omega \mid X(\omega) = x).$$

Sans risque de confusion, on note cette probabilité par $P(X = x)$.

Par exemple, si on associe la valeur 0 au résultat pair lors du lancé d'un dé équilibré à six faces, alors $X^{-1}(0) = \{2, 4, 6\}$. La probabilité d'obtenir un nombre pair est alors donnée par

$$P(\{2, 4, 6\}) = P(2) + P(4) + P(6) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}.$$

La fonction de répartition d'une variable aléatoire discrète X est définie pour tout $x \in \mathbb{R}$ par $F_X(x) = P(X \leq x)$. Plus formellement,

$$F_X(x) = P(\omega \in \Omega \mid X(\omega) \leq x).$$

On définit maintenant la moyenne d'une variable aléatoire discrète, aussi appelée *espérance*.

Définition 2.4. *L'espérance mathématique d'une variable aléatoire discrète X à valeurs dans un ensemble \mathcal{X} est définie par*

$$E(X) = \sum_{x \in \mathcal{X}} x P(X = x).$$

Voici maintenant quelques propriétés de l'espérance.

Propriété 2.2.

- (i) *Pour tout $a \in \mathbb{R}$, on a $E(X + a) = E(X) + a$;*
- (ii) *Pour tout $a \in \mathbb{R}$, on a $E(aX) = aE(X)$;*
- (iii) *Pour deux variables aléatoires discrètes X et Y , on a $E(X + Y) = E(X) + E(Y)$.*

La variance d'une variable aléatoire est également une notion très utilisée.

Définition 2.5. *La variance d'une variable aléatoire discrète est définie par*

$$\text{var}(X) = \sum_{x \in \mathcal{X}} \{x - E(X)\}^2 P(X = x).$$

À partir de la définition précédente et de la définition de l'espérance, on pourrait aussi écrire $\text{var}(X) = E\{(X - E(X))^2\}$. La variance indique le degré de dispersion d'une variable aléatoire autour de son espérance. Dans la suite, on notera $\text{var}(X) = \sigma_X^2$, où σ_X correspond à l'écart-type de X .

Voici quelques propriétés utiles concernant la variance.

Propriété 2.3.

(i) Pour tout $a \in \mathbb{R}$, on a

$$\text{var}(a) = 0, \quad \text{var}(X + a) = \text{var}(X) \quad \text{et} \quad \text{var}(aX) = a^2 \text{var}(X);$$

(ii) $\text{var}(X) = E(X^2) - \{E(X)\}^2$;

(iii) Si X et Y sont deux variables aléatoires discrètes indépendantes, alors

$$\text{var}(X + Y) = \text{var}(X) + \text{var}(Y).$$

2.2.2 Variables aléatoires continues

Contrairement à une variable aléatoire discrète, les valeurs possibles d'une variable aléatoire *continue* constituent un ensemble infini non dénombrable. En voici la définition formelle.

Définition 2.6. On appelle *variable aléatoire continue* définie sur (Ω, \mathcal{A}) , une fonction $X : \Omega \mapsto \mathbb{R}$ telle que pour tout intervalle $I \subset \mathbb{R}$,

$$X^{-1}(I) = \{\omega \in \Omega \mid X(\omega) \in I\} \in \mathcal{A}.$$

Selon ?, une variable aléatoire X est continue s'il existe une fonction non négative f telle que pour tout $B \in \mathbb{R}$,

$$P(X \in B) = \int_B f(x) dx.$$

La fonction f s'appelle la *densité* de X . Ainsi, si l'on souhaite connaître la probabilité que X prenne une valeur sur B , il suffira d'intégrer la densité de probabilité sur cet intervalle. En particulier, comme $P(X \in \mathbb{R}) = 1$, il

s'ensuit que

$$\int_{\mathbb{R}} f(x) \, dx = 1.$$

Également, pour $a \leq b \in \mathbb{R}$,

$$P(X \in [a, b]) = \int_a^b f(x) \, dx.$$

À l'instar du cas discret, la fonction de répartition d'une variable aléatoire continue X est définie par $F_X(x) = P(X \leq x)$. Cependant, comme X possède une densité de probabilité f , on peut écrire

$$F_X(x) = \int_{-\infty}^x f(s) \, ds.$$

Quelques propriétés en découlent.

Propriété 2.4. *Soit F_X , la fonction de répartition d'une certaine variable aléatoire X .*

- (i) F_X est non décroissante, c'est-à-dire que pour $x < y$, $F_X(x) \leq F_X(y)$;
- (ii) On a $\lim_{x \rightarrow -\infty} F_X(x) = 0$ et $\lim_{x \rightarrow \infty} F_X(x) = 1$;
- (iii) $P(X \in [a, b]) = F_X(b) - F_X(a)$.

L'espérance d'une variable aléatoire continue est définie par

$$E(X) = \int_{\mathbb{R}} x f(x) \, dx.$$

Cet opérateur satisfait les mêmes propriétés que celles énoncées à la Propriété 2.2 pour une variable aléatoire discrète. Aussi, la définition de la variance d'une variable continue est la même que dans le cas discret, à savoir $\text{var}(X) = E\{(X - E(X))^2\}$, et elle possède les mêmes propriétés.

2.3 Lois de probabilités discrètes

2.3.1 Distribution binomiale

Soit X , la variable aléatoire qui compte le nombre de succès obtenus lors d'une expérience aléatoire répétée n fois. Si la probabilité p d'un succès est constante au cours de n essais indépendants, alors on dit que X suit une loi Binomiale de paramètres n et p . On notera $X \sim B(n, p)$ dans ce cas.

Soit S , un *succès*, et E , un *échec*. On a alors $P(S) = p$ et $P(E) = 1 - p$. Supposons que parmi n essais, on a obtenu x succès consécutifs ; on a donc obtenu $n - x$ échecs consécutifs. Comme les essais sont indépendants, on a

$$\begin{aligned} P \left(\underbrace{S, S, \dots, S}_{x \text{ fois}}, \underbrace{E, \dots, E}_{n-x \text{ fois}} \right) &= \{P(S)\}^x \{P(E)\}^{n-x} \\ &= p^x (1-p)^{n-x}. \end{aligned}$$

Comme le nombre de façons différentes de placer x succès parmi n essais est $\binom{n}{x}$, on déduit que pour $x \in \{0, 1, \dots, n\}$,

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}.$$

Cette dernière formule, qui associe une probabilité aux différentes valeurs possibles de X , s'appelle la fonction de masse de la loi Binomiale.

2.3.2 Distribution de Bernoulli

La distribution de Bernoulli est un cas particulier de la loi Binomiale lorsque $n = 1$. On a $P(X = 0) = 1 - p$ et $P(X = 1) = p$. Si X_1, X_2, \dots, X_n sont des variables aléatoires indépendantes de loi de Bernoulli de paramètre p , alors

$$X = \sum_{i=1}^n X_i$$

suit une loi Binomiale de paramètres n et p .

2.3.3 Distribution géométrique

Soit X , la variable aléatoire qui compte le nombre d'essais nécessaires avant d'obtenir un premier succès. On dit alors que X suit une loi géométrique de paramètre p , où p est la probabilité de succès à chacun des essais. On écrit alors $X \sim \text{Geo}(p)$. Pour $x \in \mathbb{N}$, on a

$$P(X = x) = P\left(\underbrace{E, E, \dots, E}_{x-1 \text{ fois}}, S\right) = \{P(E)\}^{x-1} P(S) = p(1-p)^{x-1}.$$

En utilisant la propriété de la série géométrique, on montre facilement que

$$\sum_{x=1}^{\infty} P(X = x) = p \sum_{x=1}^{\infty} (1-p)^{x-1} = \frac{p}{1 - (1-p)} = 1.$$

De plus, la fonction de répartition de X est donnée par

$$F_X(x) = \sum_{k=1}^x P(X = k) = p \sum_{k=0}^{x-1} (1-p)^k = p \left\{ \frac{1 - (1-p)^x}{1 - (1-p)} \right\} = 1 - (1-p)^x.$$

Par conséquent, $P(X > x) = (1-p)^x$. On montre enfin que pour tout $x, y \in \mathbb{N}$, $P(X > x + y | X > x) = P(X > y)$. Cette propriété est appelée propriété de non-vieillessement de la distribution.

2.3.4 Distribution hypergéométrique

Supposons que l'on effectue n répétitions avec remise d'une expérience aléatoire à deux valeurs possibles. On peut imaginer une urne contenant un certain nombre de boules blanches et de boules rouges. Supposons que le nombre de boules dans l'urne est N , et qu'il y a R boules rouges (donc, $N - R$ boules blanches). On tire n boules avec remise et on compte le nombre X de boules rouges obtenues. Alors pour $x \in \{0, 1, \dots, n\}$,

$$P(X = x) = \frac{\binom{R}{x} \binom{N - R}{n - x}}{\binom{N}{n}}.$$

On dit que la loi de X est hypergéométrique de paramètres N , n et R , et on note $X \sim \text{Hyp}(N, n, R)$.

2.3.5 Distribution de Poisson

Supposons une variable aléatoire X_n de loi Binomiale de paramètres n et p . En posant $\lambda = np$ et en faisant tendre n vers l'infini, alors la loi de X_n converge vers une variable aléatoire X dont la fonction de masse est

$$P(X = x) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x \in \mathbb{N}.$$

On dit que X est distribuée selon la loi de Poisson de paramètre $\lambda > 0$, et on note $X \sim P(\lambda)$.

2.4 Lois de probabilités continues

2.4.1 Distribution normale

La densité de la loi Normale de paramètres $\mu \in \mathbb{R}$ et $\sigma^2 > 0$ est

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left\{ -\frac{(x - \mu)^2}{2\sigma^2} \right\}.$$

On note $X \sim N(\mu, \sigma^2)$. Si on définit $Z = (x - \mu)/\sigma$, alors on peut montrer que $Z \sim N(0, 1)$ et on l'appelle la *variable aléatoire centrée réduite*.

Cette loi apparaît dans un résultat central en statistique, à savoir le Théorème central limite. Pour le décrire, soient des variables aléatoires continues et indépendantes X_1, \dots, X_n et considérons la variable aléatoire

$$S_n = \frac{1}{n} \sum_{i=1}^n X_i.$$

Si μ est la moyenne et σ^2 est la variance de X_1, \dots, X_n , alors

$$Z_n = \sqrt{n} \left(\frac{S_n - \mu}{\sigma} \right)$$

converge en loi vers la distribution $N(0, 1)$. En utilisant ce résultat, il est possible, lorsque n est assez grand, de se servir d'une distribution Normale pour approximer la loi Binomiale. En effet, si $X \sim \text{Bin}(n, p)$, alors on a pour $x \in \mathbb{N}$ (en autant que $np \geq 5$ et $p \leq 0.5$) que

$$P(X = x) \simeq f_Y(x), \quad \text{avec } Y \sim N(np, np(1 - p)).$$

Ainsi,

$$P(X \in [a, b]) \simeq P \left(Z \leq \frac{b + 0.5 - np}{\sqrt{np(1 - p)}} \right) - P \left(Z \leq \frac{a - 0.5 - np}{\sqrt{np(1 - p)}} \right),$$

où a et b sont des entiers.

2.4.2 Distribution Gamma

Premièrement, la fonction Gamma est définie par pour $u > 0$ par

$$\Gamma(u) = \int_0^\infty x^{u-1} e^{-x} dx.$$

On peut montrer que $\Gamma(u) = (u-1)\Gamma(u-1)$ si $u > 1$. Puisque l'on trouve directement que $\Gamma(1) = 1$, alors on peut montrer que

$$\Gamma(n) = (n-1)!, \quad n \in \mathbb{N}.$$

De plus, $\Gamma(1/2) = \pi$. La densité de la loi Gamma de paramètres $\alpha > 0$ et $\beta > 0$ est donnée, pour $x > 0$, par

$$f_X(x) = \frac{x^{\alpha-1} e^{-x/\beta}}{\Gamma(\alpha) \beta^\alpha}.$$

Si on pose $\alpha = d/2$ et $\beta = 1/2$, alors on retrouve la distribution Khi-deux à d degrés de liberté. Si $\alpha = 1$ et $\beta = \lambda$, on retrouve la loi Exponentielle.

2.4.3 Distribution de Weibull

La densité de cette loi est donnée, pour $x > 0$, par

$$f_X(x) = \lambda \beta x^{\beta-1} \exp(-\lambda x^\beta).$$

En posant $\beta = 1$, on retrouve la loi Exponentielle.

2.4.4 Distribution Beta

La densité de cette loi est donnée, pour $x \in [0, 1]$, par

$$f_X(x) = \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1},$$

où $\alpha > 0$ et $\beta > 0$. On retrouve la distribution uniforme sur $(0, 1)$ lorsque $\alpha = \beta = 1$.

2.4.5 Distribution lognormale

Soit $Y \sim N(\mu, \sigma^2)$. Alors la loi de $X = e^Y$ s'appelle la distribution lognormale de paramètres μ et σ^2 . Sa densité est donnée, pour $x > 0$, par

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma x}} \exp \left\{ -\frac{(\ln(x) - \mu)^2}{2\sigma^2} \right\}.$$

2.5 Les vecteurs aléatoires

Les vecteurs aléatoires sont une généralisation des variables aléatoires. On considère ainsi un vecteur formé de d variables aléatoires.

2.5.1 Les vecteurs aléatoires discrets

Soit (X, Y) , un couple de variables aléatoires, où X et Y sont discrètes. On suppose que X prend ses valeurs dans \mathcal{X} et Y prend ses valeurs dans \mathcal{Y} . La

fonction de masse conjointe est définie pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y}$ par

$$f_{X,Y}(x, y) = P(X = x, Y = y).$$

Par définition, on a

$$f_{X,Y}(x, y) \geq 0, \quad \text{pour tout } (x, y) \in \mathcal{X} \times \mathcal{Y},$$

et

$$\sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f_{X,Y}(x, y) = 1.$$

La fonction de répartition associée à $f_{X,Y}$ est donnée par

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \sum_{s \leq x, t \leq y} f_{X,Y}(s, t).$$

On retrouve les fonctions de masse marginales de X et de Y via

$$f_X(x) = \sum_{y \in \mathcal{Y}} f_{X,Y}(x, y) \quad \text{et} \quad f_Y(y) = \sum_{x \in \mathcal{X}} f_{X,Y}(x, y).$$

Soit $A \subseteq \mathcal{X}$. Alors la fonction de masse conditionnelle de Y sachant que $X \in A$ est définie par

$$f_Y(y|X \in A) = P(Y = y|X \in A) = \frac{P(Y = y, X \in A)}{P(X \in A)}.$$

Si $A = \{x\}$, alors

$$f_Y(y|X = x) = \frac{f_{X,Y}(x, y)}{f_X(x)}.$$

On définit de la même façon la fonction de masse conditionnelle de X étant donné $Y \in B \subseteq \mathcal{Y}$.

2.5.2 Les vecteurs aléatoires continus

Soient X et Y , deux variables aléatoires continues. Tout comme dans le cas d'une variable aléatoire, le comportement stochastique de (X, Y) est caractérisé par une densité. Dans le cas à deux variables aléatoires, il s'agit d'une fonction bidimensionnelle appelée la *densité conjointe*, notée $f_{X,Y}$. On doit avoir $f_{X,Y}(x, y) \geq 0$ pour tout $(x, y) \in \mathbb{R}^2$ et

$$\int_{\mathbb{R}^2} f_{X,Y}(x, y) \, dx \, dy = 1.$$

La fonction de répartition associée à $f_{X,Y}$ est

$$F_{X,Y}(x, y) = \int_{-\infty}^x \int_{-\infty}^y f_{X,Y}(s, t) \, dt \, ds.$$

On extrait les densités marginales de X et de Y par

$$f_X(x) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dy \quad \text{et} \quad f_Y(y) = \int_{\mathbb{R}} f_{X,Y}(x, y) \, dx.$$

Soit maintenant $A \subseteq \mathbb{R}$. La densité conditionnelle de Y sachant que $X \in A$ est donnée par

$$f_Y(y|X \in A) = \frac{\int_A f_{X,Y}(x, y) \, dx}{\int_A f_X(x) \, dx}.$$

On définit de la même façon la densité conditionnelle de X sachant que $Y \in B \subseteq \mathbb{R}$.

L'espérance conditionnelle de Y sachant que $X = x$ est donnée par

$$E(Y|X = x) = \begin{cases} \sum_{y \in \mathcal{Y}} y f_Y(y|X = x) & \text{si } Y \text{ est discrète;} \\ \int_{\mathbb{R}} y f_Y(y|X = x) \, dy & \text{si } Y \text{ est continue.} \end{cases}$$

On peut montrer que

$$E(Y) = E\{E(Y|X = x)\}.$$

2.5.3 Covariance et corrélation

Soit un couple de variables aléatoires continues (X, Y) et une fonction $g : \mathbb{R}^2 \rightarrow \mathbb{R}$. On définit l'espérance mathématique de $g(X, Y)$ par

$$E\{g(X, Y)\} = \begin{cases} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} g(x, y) f_{X,Y}(x, y); \\ \int_{\mathbb{R}^2} g(x, y) f_{X,Y}(x, y) dx dy. \end{cases}$$

En particulier, si $g(x, y) = x$, alors $E\{g(X, Y)\} = E(X)$. Aussi, si $g(X, Y)$ est une combinaison linéaire de X et Y , c'est-à-dire $g(x, y) = aX + bY + c$, où $a, b, c \in \mathbb{R}$, alors $E\{g(X, Y)\} = a E(X) + b E(Y) + c$.

La covariance associée au couple (X, Y) est définie par

$$\text{cov}(X, Y) = E\{(X - E(X))(Y - E(Y))\}.$$

On peut montrer facilement que $\text{cov}(X, Y) = E(XY) - E(X)E(Y)$. Dans le cas où X et Y sont stochastiquement indépendantes, alors

$$\text{cov}(X, Y) = E(X)E(Y) - E(X)E(Y) = 0.$$

Aussi, $\text{cov}(X, X) = E(X^2) - \{E(X)\}^2 = \text{var}(X)$. Enfin, si $g(x, y) = aX + bY + c$, alors

$$\text{var}\{g(X, Y)\} = a^2 \text{var}(X) + b^2 \text{var}(Y) + 2ab \text{cov}(X, Y).$$

Le coefficient de corrélation de X et Y est donné par

$$\text{cor}(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{var}(X) \text{var}(Y)}}.$$

Celui-ci prend ses valeurs dans $[-1, 1]$. De plus, $\text{cor}(X, Y) = \pm 1$ si et seulement si la relation entre X et Y est *linéaire*, c'est-à-dire que $Y = aX + b$ pour $a, b \in \mathbb{R}$.

CHAPITRE 3

MÉTHODES DE CLASSIFICATION BAYÉSIENNE

3.1 Introduction

La classification est une méthode statistique qui a pour objectif de dégager des caractéristiques communes entre les individus d'une population. On retrouve deux méthodes principales, à savoir

- (i) l'analyse factorielle ;
- (ii) l'analyse discriminante.

L'analyse factorielle consiste à créer des axes de vecteurs représentant le mieux les variables, ce qui permet de caractériser les individus. L'analyse discriminante permet de constituer, pour les individus, des classes d'affectation dans lesquelles les caractéristiques sont similaires. La classification supervisée intervient après cette première étape. Celle-ci vise à déterminer, en s'appuyant sur des classes existantes, la classe d'appartenance d'un nouvel individu pour lequel les valeurs des variables sont connues. Dans ce chapitre, on décrira la classification Bayésienne.

3.2 Théorie

Le modèle de classification Bayésienne est un modèle probabiliste. En d'autres termes, il s'appuie sur les probabilités *a priori* des différentes classes afin de déterminer la classe d'appartenance d'une nouvelle observation.

Soient d groupes, notés $\Omega_1, \dots, \Omega_d$. Il est possible de calculer la probabilité qu'un individu appartienne au groupe Ω_k pour un certain $k \in \{1, \dots, d\}$. On note $P(x|\Omega_k)$, $x \in \mathbb{R}$, la loi de probabilité pour un individu qui appartient à la classe k . Pour déterminer à quel groupe appartient une nouvelle observation dont les caractéristiques sont $x \in \mathbb{R}$, il faut connaître $P(\Omega_k|x)$ pour chaque $k \in \{1, \dots, d\}$, à savoir la probabilité d'appartenir à Ω_k sachant que x s'est réalisé. Du Théorème de Bayes,

$$P(\Omega_k|x) = \frac{P(x|\Omega_k) P(\Omega_k)}{P(x)},$$

où $P(\Omega_k)$ est la probabilité *a priori* de la classe k et

$$P(x) = \sum_{k=1}^d P(x|\Omega_k) P(\Omega_k).$$

On a donc

$$P(\Omega_k|x) = \frac{P(x|\Omega_k) P(\Omega_k)}{\sum_{k=1}^d P(x|\Omega_k) P(\Omega_k)}.$$

Puisque le dénominateur de cette dernière expression est le même quel que soit $k \in \{1, \dots, d\}$,

$$P(\Omega_k|x) \propto P(x|\Omega_k) P(\Omega_k).$$

Par conséquent, la classe d'appartenance pour un individu dont la valeur observée est x sera

$$K = \arg \max_{k \in \{1, \dots, d\}} P(x|\Omega_k) P(\Omega_k).$$

Dans le cas où les caractéristiques des individus sont des vecteurs dans \mathbb{R}^p , la classification Bayésienne est, à la base, fondée sur le même principe que dans le cas univarié. Ainsi, si $\mathbf{x} \in \mathbb{R}^p$ correspond aux valeurs prises par un nouvel individu, alors la probabilité d'appartenance à une classe Ω_k , $k \in \{1, \dots, d\}$, est donnée par

$$P(\Omega_k|\mathbf{x}) = \frac{P(\mathbf{x}|\Omega_k) P(\Omega_k)}{\sum_{k=1}^d P(\mathbf{x}|\Omega_k) P(\Omega_k)}.$$

Par les mêmes arguments que pour le cas univarié, la classe d'appartenance pour un individu dont la valeur observée est \mathbf{x} sera

$$K = \arg \max_{k \in \{1, \dots, d\}} P(\mathbf{x}|\Omega_k) P(\Omega_k).$$

Afin de faciliter les calculs subséquents, posons

$$g_k(\mathbf{x}) = \ln \{P(\mathbf{x}|\Omega_k) P(\Omega_k)\} = \ln P(\mathbf{x}|\Omega_k) + \ln P(\Omega_k).$$

On appellera g_1, \dots, g_d les *fonctions discriminantes*. À noter que puisque la fonction logarithmique est monotone croissante,

$$K = \arg \max_{k \in \{1, \dots, d\}} g_k(\mathbf{x}).$$

Souvent, on émet l'hypothèse de la *normalité des observations*, c'est-à-dire que l'on suppose que $P(\mathbf{x}|\Omega_k)$ est la loi Normale p -dimensionnelle de moyenne $\mu_k \in \mathbb{R}^p$ et de matrice de variance-covariance $\Sigma_k \in \mathbb{R}^{p \times p}$. Dans la suite, on

notera cette loi $\mathbb{N}_p(\mu_k, \Sigma_k)$. On peut alors montrer que les fonctions discriminantes s'écrivent

$$g_k(\mathbf{x}) = -\frac{1}{2} (\mathbf{x} - \mu_k)^\top \Sigma_k^{-1} (\mathbf{x} - \mu_k) - \frac{p}{2} \ln(2\pi) - \frac{1}{2} \ln |\Sigma_k| + \ln P(\Omega_k).$$

Quelques cas particuliers décrits par ? seront maintenant examinés.

3.3 Loi Normale sous l'indépendance

Supposons que pour chaque $k \in \{1, \dots, d\}$, la matrice de variance-covariance s'écrit $\Sigma_k = \sigma^2 I_p$, où I_p est la matrice identité de dimension $p \times p$. Ce cas se produit si les p caractéristiques sont indépendantes et possèdent la même variance σ^2 . En négligeant les termes constants, on obtient alors

$$g_k(\mathbf{x}) = -\frac{\|\mathbf{x} - \mu_k\|^2}{2\sigma^2} + \ln P(\Omega_k),$$

où pour $\mathbf{a} = (a_1, \dots, a_p) \in \mathbb{R}^p$,

$$\|\mathbf{a}\| = \sqrt{a_1^2 + \dots + a_p^2}$$

est la norme euclidienne. En développant l'expression pour la norme, on obtient finalement

$$g_k(\mathbf{x}) = -\frac{1}{2\sigma^2} (\mathbf{x}^\top \mathbf{x} - 2\mu_k^\top \mathbf{x} + \mu_k^\top \mu_k) + \ln P(\Omega_k).$$

En omettant les termes qui sont communs à tous les $k \in \{1, \dots, d\}$, on déduit des fonctions discriminantes linéaires de la forme

$$g_k(\mathbf{x}) = \mathbf{w}_k^\top \mathbf{x} + w_{k0},$$

où

$$\mathbf{w}_k = \frac{1}{\sigma^2} \mu_k \quad \text{et} \quad w_{k0} = -\frac{1}{2\sigma^2} \mu_k^\top \mu_k + \ln P(\Omega_k).$$

Graphiquement, le seuil de décision pour un modèle décisionnel linéaire est déterminé par le plan défini par l'équation $g_k(\mathbf{x}) = g_{k'}(\mathbf{x})$, où k et k' correspondent aux deux classes pour lesquelles les probabilités *a priori* sont les plus élevées. Pour ce cas en particulier, cela revient à résoudre l'équation

$$\mathbf{w}^\top (\mathbf{x} - \mathbf{x}_0) = 0,$$

où $\mathbf{w} = \mu_k - \mu_{k'}$ et

$$\mathbf{x}_0 = \frac{1}{2} (\mu_k + \mu_{k'}) - \frac{\sigma^2}{\|\mu_k - \mu_{k'}\|} \ln \frac{P(\Omega_k)}{P(\Omega_{k'})} (\mu_k - \mu_{k'}).$$

Si les probabilités *a priori* sont égales pour toutes les classes, les termes $\ln P(\Omega_1), \dots, \ln P(\Omega_d)$ peuvent être ignorés et la règle de décision devient plus simple. Ainsi, pour classer une nouvelle observation dont la valeur est \mathbf{x} , il suffit de choisir la classe pour laquelle $\|\mathbf{x} - \mu_k\|$ est minimale. Donc, la classe sélectionnée est

$$K = \arg \min_{k \in \{1, \dots, d\}} \|\mathbf{x} - \mu_k\|.$$

3.4 Loi Normale avec matrices de variance-covariance homogènes

Supposons que $\Sigma_1 = \dots = \Sigma_d = \Sigma$, c'est-à-dire que les matrices de variance-covariance pour chacune des classes sont identiques. En ignorant les termes

constants, on montre que dans ce cas,

$$g_k(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_k)^\top \Sigma^{-1}(\mathbf{x} - \mu_k) + \ln P(\Omega_k).$$

En développant le terme $(\mathbf{x} - \mu_k)^\top$ et en ne conservant que les termes non constants, on peut écrire

$$g_k(\mathbf{x}) = \mathbf{w}_k^\top \mathbf{x} + w_{i0},$$

où

$$\mathbf{w}_k = \Sigma^{-1} \mu_k \quad \text{et} \quad w_{i0} = -\frac{1}{2} \mu_k^\top \Sigma^{-1} \mu_k + \ln P(\Omega_k).$$

Tout comme dans le cas de composantes indépendantes décrit à la section 3.3, la règle de décision consistera à résoudre

$$\mathbf{w}^\top (\mathbf{x} - \mathbf{x}_0) = 0,$$

où $\mathbf{w} = \Sigma^{-1}(\mu_k - \mu_{k'})$ et

$$\mathbf{x}_0 = \frac{1}{2}(\mu_k + \mu_{k'}) - \frac{\ln P(\Omega_k)}{\ln P(\Omega_{k'})} \frac{(\mu_k - \mu_{k'})}{(\mu_k - \mu_{k'})^\top \Sigma^{-1}(\mu_k - \mu_{k'})}.$$

Pour des probabilités *a priori* supposées égales, c'est-à-dire $P(\Omega_1) = \dots = P(\Omega_d)$, la règle de décision consiste à assigner une nouvelle observation dont la valeur est \mathbf{x} dans la classe telle que la distance de Mahalanobis, à savoir

$$(\mathbf{x} - \mu_k)^\top \Sigma^{-1}(\mathbf{x} - \mu_k),$$

est minimale. Ainsi, la classe sélectionnée est

$$K = \arg \min_{k \in \{1, \dots, d\}} (\mathbf{x} - \mu_k)^\top \Sigma^{-1}(\mathbf{x} - \mu_k).$$

En posant $\Sigma = \sigma^2 I_p$, on retrouve

$$\begin{aligned} K &= \arg \min_{k \in \{1, \dots, d\}} \sigma^2 (\mathbf{x} - \mu_k)^\top (\mathbf{x} - \mu_k) \\ &= \arg \min_{k \in \{1, \dots, d\}} \sigma^2 \|\mathbf{x} - \mu_k\|^2 \\ &= \arg \min_{k \in \{1, \dots, d\}} \|\mathbf{x} - \mu_k\|, \end{aligned}$$

tel que mentionné à la section 3.3.

3.5 Loi Normale avec matrices de variance-covariance hétérogènes

En n'émettant aucune hypothèse sur la forme des matrices de variance-covariance, on a

$$g_k(\mathbf{x}) = \mathbf{x}^\top \mathbf{W}_k \mathbf{x} + \mathbf{w}_k^\top \mathbf{x} + w_{i0},$$

où $\mathbf{W}_k = \Sigma_k^{-1}/2$, $\mathbf{w}_k = \Sigma_k^{-1} \mu_k$ et

$$w_{i0} = -\frac{1}{2} \mu_k^\top \Sigma_k^{-1} \mu_k - \frac{1}{2} \ln |\Sigma_k| + \ln P(\Omega_k).$$

Un exemple de calcul est donné dans la suite.

Exemple 3.1. Soient $d = 2$ classes d'une population à $p = 2$ dimensions.

Les moyennes et les matrices de variance-covariance sont respectivement

$$\mu_1 = \begin{bmatrix} 3 \\ 6 \end{bmatrix}, \quad \Sigma_1 = \begin{pmatrix} 1/2 & 0 \\ 0 & 2 \end{pmatrix}$$

et

$$\mu_2 = \begin{bmatrix} 3 \\ -2 \end{bmatrix}, \quad \Sigma_2 = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}.$$

De là, on obtient que les matrices de variance-covariance inverses sont

$$\Sigma_1^{-1} = \begin{pmatrix} 2 & 0 \\ 0 & 1/2 \end{pmatrix} \quad \text{et} \quad \Sigma_2^{-1} = \begin{pmatrix} 1/2 & 0 \\ 0 & 1/2 \end{pmatrix}.$$

Supposons un individu dont la valeur observée est $\mathbf{x} = (x_1, x_2)$. À partir des formules précédentes, on obtient que les fonctions discriminantes sont

$$g_1(\mathbf{x}) = -x_1^2 - \frac{x_2^2}{4} + 6x_1 + 3x_2 - 18.693$$

et

$$g_2(\mathbf{x}) = -\frac{x_1^2}{4} - \frac{x_2^2}{4} + \frac{3x_1}{2} - x_2 - 4.636.$$

Afin d'obtenir la règle de décision, on doit résoudre $g_1(\mathbf{x}) - g_2(\mathbf{x}) = 0$. On déduit alors l'équation de second degré

$$x_2 = \frac{3}{16}x_1^2 - \frac{9}{8}x_1 + 3.514.$$

Cette courbe est présentée à la Figure 3.1. Ainsi un individu, dont les coordonnées $\mathbf{x} = (x_1, x_2)$ sont au-dessus de cette courbe, sera assigné à la classe Ω_1 car alors, $g_1(\mathbf{x}) > g_2(\mathbf{x})$. Dans le cas contraire, il sera assigné à la classe Ω_2 puisque $g_2(\mathbf{x}) > g_1(\mathbf{x})$.

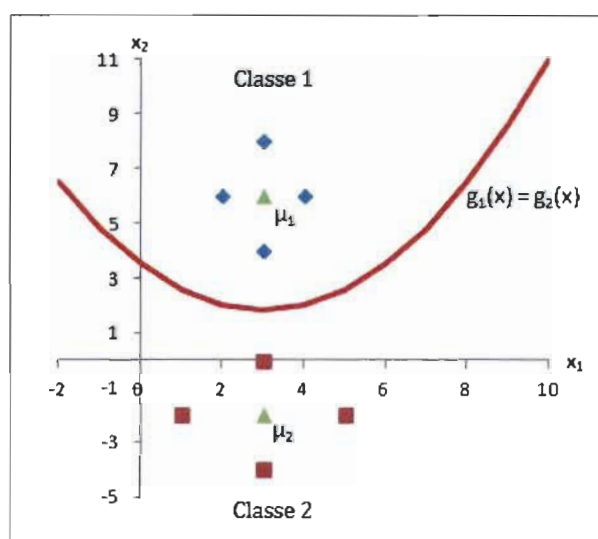


FIGURE 3.1: Exemple de règle de décision bayésienne

CHAPITRE 4

COPULES : THÉORIE ET ESTIMATION

La notion de *copule*, instaurée par ?, est un outil probabiliste permettant d'extraire la dépendance qui existe entre plusieurs variables aléatoires. C'est donc un concept qui ne tient pas compte des comportements marginaux des variables dans un vecteur aléatoire. Un des avantages majeurs de cette théorie est la possibilité de construire des modèles multivariés avec les lois marginales désirées.

4.1 Définition et Théorème de Sklar

Le cas à deux variables sera traité dans la suite. L'extension multidimensionnelle est relativement directe et sera brièvement présentée à la fin de cette section. Soit donc (X, Y) , un couple de variables aléatoires de fonction de répartition conjointe $H(x, y) = P(X \leq x, Y \leq y)$, où $(x, y) \in \mathbb{R}^2$. Les lois marginales sont données par

$$F(x) = P(X \leq x) = \lim_{y \rightarrow \infty} H(x, y) \quad \text{et} \quad G(y) = P(Y \leq y) = \lim_{x \rightarrow \infty} H(x, y).$$

Le fameux Théorème de ? peut maintenant être énoncé.

Théorème 4.1. *Soit H , une fonction de répartition bivariée de marges continues F et G . Alors il existe une unique copule $C : [0, 1]^2 \rightarrow [0, 1]$ telle que*

$$H(x, y) = C \{F(x), G(y)\}. \quad (4.1)$$

La copule C correspond à la loi conjointe de $(F(X), G(Y))$. En effet,

$$\begin{aligned} P \{F(X) \leq u, G(Y) \leq v\} &= P \{X \leq F^{-1}(u), Y \leq G^{-1}(v)\} \\ &= H \{F^{-1}(u), G^{-1}(v)\} \\ &= C \{F \circ F^{-1}(u), G \circ G^{-1}(v)\} \\ &= C(u, v). \end{aligned}$$

Puisque $F(X)$ et $G(Y)$ sont uniformément distribuées sur $(0, 1)$, on voit que C est une fonction de répartition sur $[0, 1]^2$ dont les marges sont uniformes.

Le Théorème de ? met en lumière le fait que la loi conjointe de (X, Y) est composée de trois éléments, à savoir (i) le comportement marginal de X (représenté par F), (ii) le comportement marginal de Y (représenté par G), ainsi que (iii) la dépendance entre X et Y (représentée par C).

Pour qu'une fonction $C : [0, 1]^2 \rightarrow [0, 1]$ soit la copule d'une certaine loi bivariée elle doit satisfaire certaines conditions. Ainsi, pour tout $u \in [0, 1]$, on doit avoir

$$C(u, 0) = C(0, u) = 0 \quad \text{et} \quad C(u, 1) = C(1, u) = u.$$

De plus, C doit être 2-croissante, c'est-à-dire que pour tout $u_1 \leq u_2$ et $v_1 \leq v_2$,

$$C(u_2, v_2) - C(u_1, v_2) - C(u_2, v_1) + C(u_1, v_1) \geq 0.$$

Ainsi à partir d'une fonction C , satisfaisant les conditions pour être une copule, on peut bâtir une loi bivariée sur \mathbb{R}^2 (ou encore sur un sous-ensemble de \mathbb{R}^2) avec des marges données. L'exemple suivant en donne une illustration.

Exemple 4.1. *On peut montrer que la fonction*

$$C(u, v) = \frac{uv}{u + v - uv}.$$

est une copule. En prenant des marges exponentielles, à savoir $F(x) = G(x) = 1 - e^{-x}$, on construit le modèle

$$\begin{aligned} H(x, y) &= C\{F(x), G(y)\} \\ &= \frac{(1 - e^{-x})(1 - e^{-y})}{(1 - e^{-x}) + (1 - e^{-y}) - (1 - e^{-x})(1 - e^{-y})} \\ &= \left(\frac{1}{1 - e^{-x}} + \frac{1}{1 - e^{-y}} - 1 \right)^{-1}. \end{aligned}$$

La fonction H est donc une fonction de répartition sur $[0, \infty) \times [0, \infty)$ dont la dépendance est représentée par C et dont les marges sont exponentielles.

Une conséquence du Théorème de ? est la possibilité d'*extraire* la copule qui est sous-jacente à une loi bivariée. Ainsi, en posant $u = F(x)$ et $v = G(y)$ dans l'équation (4.1), on déduit

$$C(u, v) = H\{F^{-1}(u), G^{-1}(v)\}. \quad (4.2)$$

Une illustration est donnée dans l'exemple suivant.

Exemple 4.2. *Soit la fonction de répartition sur $[0, 1]^2$ définie par*

$$H(x, y) = \sqrt{\frac{xy}{x + y - xy}}.$$

Ses marges sont $F(x) = H(x, 1) = \sqrt{x}$ et $G(y) = H(1, y) = \sqrt{y}$. On a donc facilement $F^{-1}(u) = u^2$ et $G^{-1}(v) = v^2$. Une application directe de l'équation (4.2) amène

$$C(u, v) = H(u^2, v^2) \sqrt{\frac{u^2 v^2}{u^2 + v^2 - u^2 v^2}} = (u^{-2} + v^{-2} - 1)^{-1/2}.$$

Il s'agit d'un cas particulier de la copule de Clayton (voir ?), qui sera décrite plus loin dans le mémoire.

Afin de généraliser le Théorème de ? au cas à d dimensions, soit un vecteur aléatoire $\mathbf{X} = (X_1, \dots, X_d)$ de loi conjointe

$$H(x_1, \dots, x_d) = P(X_1 \leq x_1, \dots, X_d \leq x_d).$$

Si $F_j(x) = P(X_j \leq x)$, $j \in \{1, \dots, d\}$, sont les marges de H et qu'elles sont continues, alors il existe une unique fonction $C : [0, 1]^d \rightarrow [0, 1]$ telle que

$$H(x_1, \dots, x_d) = C\{F_1(x_1), \dots, F_d(x_d)\}.$$

4.2 Propriétés des copules

4.2.1 Invariance d'une copule

La copule d'un couple de variables aléatoires (X, Y) est invariante sous des transformations strictement croissantes des marges. En d'autres termes, la copule de $(\kappa(X), \eta(Y))$, où κ et η sont des fonctions monotones croissantes,

est la même que la copule C de (X, Y) . En effet, la loi de $(\kappa(X), \eta(Y))$ est

$$\begin{aligned} \mathbb{P} \{ \kappa(X) \leq x, \eta(Y) \leq y \} &= \mathbb{P} \{ X \leq \kappa^{-1}(x), Y \leq \eta^{-1}(y) \} \\ &= H \{ \kappa^{-1}(x), \eta^{-1}(y) \} \\ &= C \{ F \circ \kappa^{-1}(x), G \circ \eta^{-1}(y) \}. \end{aligned}$$

Puisque les lois marginales de $\kappa(X)$ et de $\eta(Y)$ sont respectivement $F \circ \kappa^{-1}(x)$ et $G \circ \eta^{-1}(y)$, on déduit que la copule de $(\kappa(X), \eta(Y))$ est C .

4.2.2 Copule d'indépendance et bornes de Fréchet

On sait que deux variables aléatoires X et Y sont indépendantes si et seulement si leur loi conjointe H se factorise, c'est-à-dire si

$$H(x, y) = F(x) \times G(y),$$

où $F(x) = \mathbb{P}(X \leq x)$ et $G(y) = \mathbb{P}(Y \leq y)$. Par une application de l'équation (4.2), la copule associée à cette situation, dénommée la copule d'indépendance, est

$$\Pi(u, v) = H \{ F^{-1}(u), G^{-1}(v) \} = u \times v.$$

Il existe des copules *minimale* et *maximale*, appelées les bornes de Fréchet-Hoeffding, telles que toute copule C est bornée inférieurement et supérieurement. En effet, on peut montrer que pour tout $(u, v) \in [0, 1]^2$,

$$W(u, v) \leq C(u, v) \leq M(u, v)$$

où

$$W(u, v) = \max(u + v - 1, 0) \quad \text{et} \quad M(u, v) = \min(u, v).$$

La copule $M(u, v)$ est associée à la *dépendance positive parfaite* ; elle correspond à la loi de (U, V) quand $V = U$, où U est uniforme sur $(0, 1)$. La copule W est associée à la *dépendance négative parfaite* et correspond à la loi de (U, V) lorsque $V = 1 - U$.

4.3 Exemples de copules

4.3.1 Copule Normale

Soit ϕ_ρ , la densité de la loi Normale bivariée standard de coefficient de corrélation ρ . Alors la fonction de répartition associée est

$$\Phi_\rho(x, y) = \int_{-\infty}^x \int_{-\infty}^y h_\rho(s, t) dt ds.$$

Soit maintenant Φ la fonction de répartition de la loi $\mathbb{N}(0, 1)$. Comme les marges de Φ_ρ sont Φ , alors une application du Théorème de Sklar inverse permet de déduire que la copule Normale est donnée implicitement par

$$C_\rho^N(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} h_\rho(s, t) dt ds.$$

Les densités de la copule Normale lorsque $\rho = .5$ et $\rho = .9$ sont présentées en haut de la Figure 4.1. Les courbes de niveau correspondantes sont présentées dans le bas de la figure.

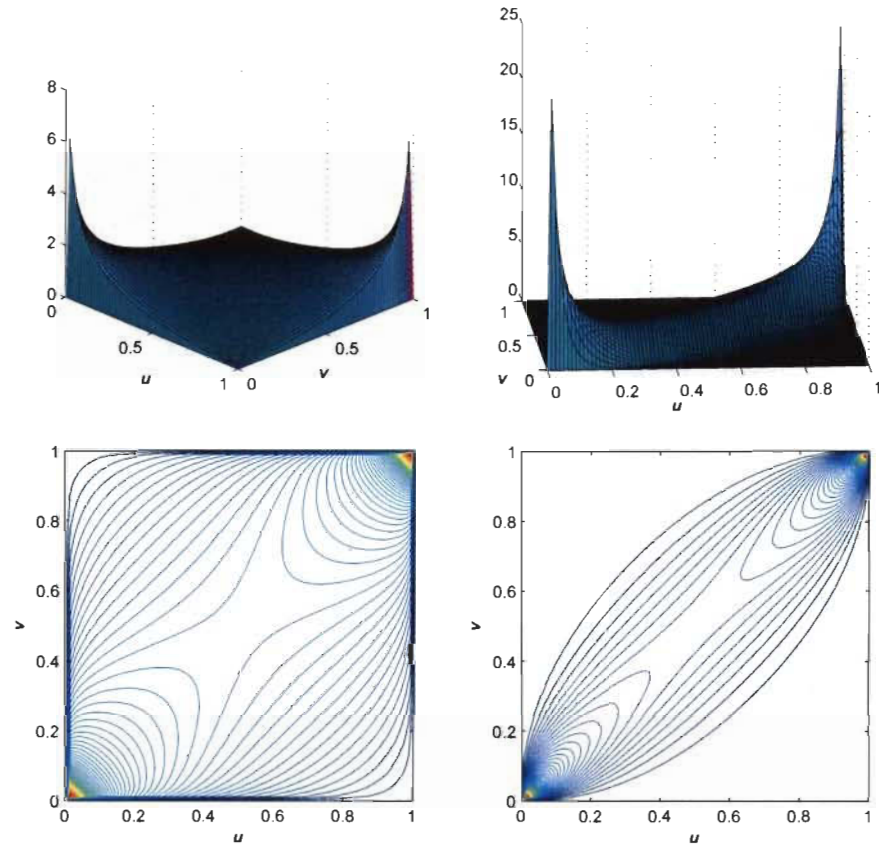


FIGURE 4.1: En haut : densités de la copule Normale quand $\rho = .5$ (à gauche) et $\rho = .9$ (à droite) ; en bas : courbes de niveau de la copule Normale quand $\rho = .5$ (à gauche) et $\rho = .9$ (à droite)

4.3.2 Copule de Student

La copule de Student se déduit de façon similaire à la copule Normale. Soit donc $T_{\rho, \nu}$, la fonction de répartition de la loi de Student bivariée à ν degrés de liberté et un coefficient de corrélation ρ . Comme les marges de cette loi

sont la distribution de Student à ν degrés de liberté, notée T_ν , on obtient que la copule de Student est

$$C_{\rho,\nu}(u,v) = T_{\rho,\nu} \{T_\nu^{-1}(u), T_\nu^{-1}(v)\}.$$

Une illustration pour ce modèle quand $\nu = 1$ et $\rho = .5$ est présentée à la Figure 4.2.

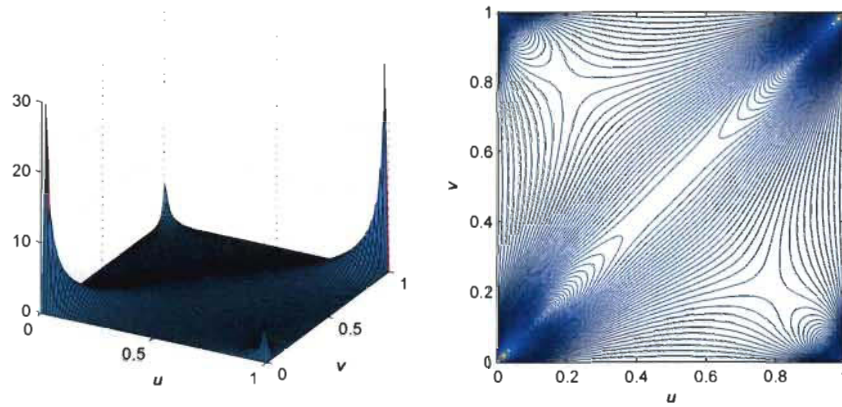


FIGURE 4.2: Densité (à gauche) et courbes de niveau (à droite) de la copule de Student à 1 degré de liberté et $\rho = .5$

4.3.3 Copules Archimédiennes

Une copule C fait partie de la famille Archimédienne si elle peut s'écrire sous la forme

$$C(u,v) = \varphi^{-1} \{ \varphi(u) + \varphi(v) \},$$

où $\varphi : [0, 1] \mapsto [0, \infty)$ est un générateur décroissant et convexe tel que $\varphi(1) = 0$. Ces modèles ont été étudiés par ?. La généralisation à d dimensions est

$$C(u_1, \dots, u_d) = \varphi^{-1} \{ \varphi(u_1) + \dots + \varphi(u_d) \},$$

où $\varphi(1) = 0$ et pour tout $k \in \{1, \dots, d\}$,

$$(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) > 0.$$

Le *Critère d'Abel* mentionne qu'une copule est Archimédienne si et seulement si ses dérivées partielles

$$C_{10}(u, v) = \frac{\partial}{\partial u} C(u, v) \quad \text{et} \quad C_{01}(u, v) = \frac{\partial}{\partial v} C(u, v)$$

existent et s'il existe une fonction intégrable $\xi : (0, 1) \rightarrow (0, \infty)$ telle que

$$\xi(u) C_{01}(u, v) = \xi(v) C_{10}(u, v)$$

pour tout $(u, v) \in [0, 1]^2$.

Une copule Archimédienne bivariées C_φ de générateur φ est associative au sens où pour tout $u, v, w \in [0, 1]$,

$$C_\varphi \{u, C_\varphi(v, w)\} = C_\varphi \{C_\varphi(u, v), w\}.$$

En effet, par des calculs directs, on a

$$\begin{aligned} C_\varphi \{u, C_\varphi(v, w)\} &= \varphi^{-1} [\varphi(u) + \varphi \{ \varphi^{-1}(\varphi(v) + \varphi(w)) \}] \\ &= \varphi^{-1} \{ \varphi(u) + \varphi(v) + \varphi(w) \} \end{aligned}$$

et

$$\begin{aligned} C_\varphi \{C_\varphi(u, v), w\} &= \varphi^{-1} [\varphi \{ \varphi^{-1}(\varphi(u) + \varphi(v)) \} + \varphi(w)] \\ &= \varphi^{-1} \{ \varphi(u) + \varphi(v) + \varphi(w) \}. \end{aligned}$$

Il est également facile à montrer qu'une copule Archimédienne est symétrique, c'est-à-dire que $C_\varphi(u, v) = C_\varphi(v, u)$ pour tout $(u, v) \in [0, 1]^2$. En effet,

$$C_\varphi(u, v) = \varphi^{-1} \{ \varphi(u) + \varphi(v) \} = \varphi^{-1} \{ \varphi(v) + \varphi(u) \} = C_\varphi(v, u).$$

Supposons maintenant une constante positive A et soit le générateur $\tilde{\varphi}(t) = A \varphi(t)$ défini à partir d'un générateur Archimédien φ . Alors $\tilde{\varphi}$ induit la même copule Archimédienne que φ . En effet, puisque $\tilde{\varphi}^{-1}(x) = \varphi^{-1}(x/A)$, on a

$$\begin{aligned} C_{\tilde{\varphi}}(u, v) &= \tilde{\varphi}^{-1} \{ \tilde{\varphi}(u) + \tilde{\varphi}(v) \} \\ &= \varphi^{-1} \left\{ \frac{1}{A} (A \varphi(u) + A \varphi(v)) \right\} \\ &= \varphi^{-1} \{ \varphi(u) + \varphi(v) \} \\ &= C_\varphi(u, v). \end{aligned}$$

Un cas particulier survient lorsque $\varphi(t) = -\ln t$. Dans ce cas, $\varphi^{-1} = \exp(-t)$ et la copule correspondante est

$$C(u, v) = \exp[-\{(-\ln u) + (-\ln v)\}] = uv,$$

à savoir la copule d'indépendance. Dans la suite, les populaires modèles de copules de Clayton et de Frank sont décrits.

Exemple 4.3. *La copule de Clayton est définie par*

$$C_\theta^{\text{CL}}(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta},$$

où $\theta > 0$. À la limite quand $\theta \rightarrow 0$, on retrouve la copule d'indépendance. Le générateur de cette copule est

$$\varphi_\theta^{\text{CL}}(t) = \frac{t^{-\theta} - 1}{\theta}.$$

Les densités de cette copule lorsque $\theta = 1$ et $\theta = 3$ sont présentées à la Figure 4.3.

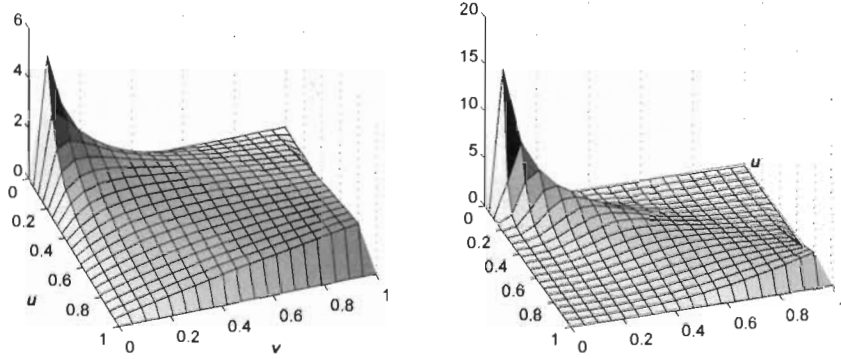


FIGURE 4.3: Densités de la copule de Clayton pour $\theta = 1$ (à gauche) et $\theta = 3$ (à droite)

Exemple 4.4. La copule de Frank est définie pour $\theta \in \mathbb{R} \setminus \{0\}$ par

$$C_{\theta}^F(u, v) = -\frac{1}{\theta} \ln \left\{ 1 - \frac{(1 - e^{-\theta u})(1 - e^{-\theta v})}{1 - e^{-\theta}} \right\}.$$

Le générateur quant à lui s'écrit de la manière suivante

$$\varphi_{\theta}^F(t) = \log \left(\frac{1 - e^{-\theta}}{1 - e^{-\theta t}} \right).$$

Les densités de cette copule lorsque $\theta = 1$ et $\theta = 3$ sont présentées à la Figure 4.4.

4.4 Mesures de dépendance

Le coefficient de corrélation présenté au Chapitre 2 permet de mesurer le degré de dépendance entre deux variables aléatoires X et Y . Cependant,

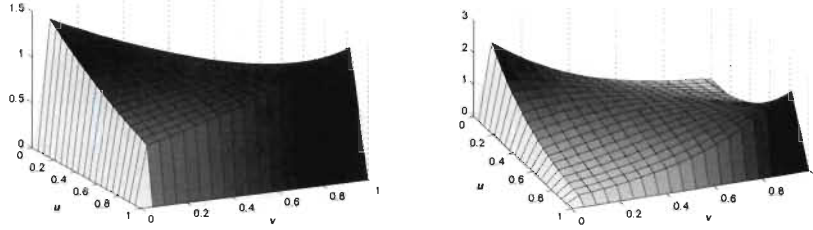


FIGURE 4.4: Densités de la copule de Frank pour $\theta = 1$ (à gauche) et $\theta = 3$ (à droite)

cette mesure présente plusieurs inconvénients. Notamment, elle n'est pas invariante sous des transformations monotones des marges. Les mesures de dépendances telles le tau de Kendall et le rho de Spearman ne souffrent pas de cette limitation. Ainsi, on acquiert une information plus fiable concernant la dépendance entre deux variables aléatoires.

4.4.1 Tau de Kendall

La définition du tau de Kendall est basée sur la notion de concordance entre deux couples de variables aléatoires. Soient donc deux couples aléatoires indépendants (X_1, Y_1) et (X_2, Y_2) , tous deux de loi conjointe H . On dit que ces couples sont concordants si

$$X_1 < X_2 \text{ et } Y_1 < Y_2 \quad \text{ou si} \quad X_1 > X_2 \text{ et } Y_1 > Y_2.$$

A l'inverse, on dit qu'ils sont discordants si

$$X_1 < X_2 \text{ et } Y_1 > Y_2 \quad \text{ou si} \quad X_1 > X_2 \text{ et } Y_1 < Y_2.$$

Le tau de Kendall est défini comme la différence entre la probabilité de concordance et la probabilité de discordance. Spécifiquement, si \mathcal{C} représente l'événement *concordance* et \mathcal{D} l'événement *discordance*, alors le tau de Kendall est défini par

$$\tau_H = P(\mathcal{C}) - P(\mathcal{D}).$$

Cette mesure de dépendance ne dépend en fait que de l'unique copule que l'on peut extraire de la loi H via le Théorème de ?. En effet, on a

$$\begin{aligned} P(\mathcal{C}) &= P(X_1 < X_2, Y_1 < Y_2) + P(X_2 < X_1, Y_2 < Y_1) \\ &= 2 P(X_1 < X_2, Y_1 < Y_2) \\ &= 2 P\{F(X_1) < F(X_2), G(Y_1) < G(Y_2)\} \\ &= \int_0^1 \int_0^1 C(u, v) dC(u, v). \end{aligned}$$

Ainsi, comme $P(\mathcal{D}) = 1 - P(\mathcal{C})$,

$$\tau_H = \tau_C = 4 \int_0^1 \int_0^1 C(u, v) dC(u, v) - 1.$$

Pour une copule Archimédienne d'un générateur φ , on a

$$\tau_C = 1 + 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt,$$

alors que pour la copule Normale,

$$\tau_C = \frac{2}{\pi} \sin^{-1}(\rho).$$

Enfin, on peut estimer τ_C à partir d'observations i.i.d. $(X_1, Y_1), \dots, (X_n, Y_n)$.

Soient \mathcal{C}_n et \mathcal{D}_n , respectivement le nombre de concordances et le nombre de discordances parmi les $n(n-1)/2$ comparaisons de couples possibles. Alors une estimation sans biais de τ_C est

$$\tau_n = \mathcal{C}_n - \mathcal{D}_n \Big/ \binom{n}{2}.$$

4.4.2 Rho de Spearman

Soient des observations i.i.d. $(X_1, Y_1), \dots, (X_n, Y_n)$. Une mesure de dépendance introduite par ? est le coefficient de corrélation de rangs. Pour la définir, soit R_i , le rang de X_i parmi X_1, \dots, X_n et S_i , le rang de Y_i parmi Y_1, \dots, Y_n . Alors le rho de Spearman empirique est

$$r_S = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^n (R_i - S_i)^2.$$

Puisque les rangs sont invariants suite à une transformation monotone croissante des observations, la mesure r_S est également invariante.

La version théorique de r_S , notée ρ_S , est basée sur la corrélation usuelle mais elle est calculée à partir de $F(X)$ et $G(Y)$, où $(X, Y) \sim H$. Ainsi,

$$\rho_S = \frac{\mathbb{E}\{F(X)G(Y)\} - \mathbb{E}\{F(X)\}\mathbb{E}\{G(Y)\}}{\sqrt{\text{var}\{F(X)\}\text{var}\{G(Y)\}}}.$$

Puisque $F(X)$ et $G(Y)$ sont des variables uniformes sur $(0, 1)$, on a

$$\mathbb{E}\{F(X)\} = \mathbb{E}\{G(Y)\} = \frac{1}{2} \quad \text{et} \quad \text{var}\{F(X)\} = \text{var}\{G(Y)\} = \frac{1}{12},$$

ce qui amène la simplification $\rho_S = 12 \mathbb{E}\{F(X)G(Y)\} - 3$. Enfin, puisque $(F(X), G(Y)) \sim C$, où C est la copule de H ,

$$\rho_S = 12 \int_0^1 \int_0^1 uv \, dC(u, v) - 3.$$

De cette façon, à l'instar du tau de Kendall, la mesure de dépendance de Spearman ne dépend que de la copule sous-jacente à une loi bidimensionnelle.

CHAPITRE 5

UNE MÉTHODE DE CLASSIFICATION À BASE DE COPULES

5.1 Objectif

On a vu au Chapitre 3 que la classification Bayésienne est fondée en général sur la loi Normale multidimensionnelle. C'est donc sur cette base que sont obtenues les fonctions discriminantes qui servent à la classification. Cependant, l'hypothèse de normalité des données ne peut pas toujours être soutenue de façon raisonnable. Autrement dit, il arrive que la distribution des observations n'appartienne pas à la famille des lois Normales, ce qui veut dire que les conclusions d'une classification *classique* peuvent être erronées.

L'objectif de ce chapitre est de pallier ces éventuels écarts à la normalité en proposant une alternative aux méthodes bayésiennes classiques fondée sur les copules. L'idée principale consiste à établir une procédure de classification basée sur une loi générale construite à partir d'une certaine copule paramétrique. Une attention particulière sera portée à des modèles issus de la famille des copules Archimédiennes.

5.2 Méthodologie

Soit une population d -dimensionnelle partitionnée en classes $\Omega_1, \dots, \Omega_K$. Supposons une nouvelle observation $\mathbf{x} \in \mathbb{R}^d$ que l'on souhaite affecter à une des classes de manière optimale. Pour des copules C_1, \dots, C_K , la fonction de répartition associée à la classe $k \in \{1, \dots, K\}$ s'écrit

$$F_k(\mathbf{x}) = C_k \{F_{1k}(x_1), \dots, F_{dk}(x_d)\},$$

où F_{1k}, \dots, F_{dk} sont les fonctions de répartition marginales pour la classe k . Si f_{1k}, \dots, f_{dk} sont les densités de probabilité associées à F_{1k}, \dots, F_{dk} , alors la densité de la classe k est

$$\begin{aligned} P(\mathbf{x}|\Omega_k) &= \frac{\partial^d}{\partial x_1 \dots \partial x_d} F_k(\mathbf{x}) \\ &= \frac{\partial^d}{\partial x_1 \dots \partial x_d} C_k \{F_{1k}(x_1), \dots, F_{dk}(x_d)\} \\ &= c_k \{F_{1k}(x_1), \dots, F_{dk}(x_d)\} \prod_{j=1}^d f_{jk}(x_j), \end{aligned} \quad (5.1)$$

où c_k est la densité associée à C_k . En définissant maintenant pour chaque $k \in \{1, \dots, K\}$ les fonctions discriminantes $g_k(\mathbf{x}) = \ln P(\mathbf{x}|\Omega_k)$, on a

$$g_k(\mathbf{x}) = \ln c_k \{F_{1k}(x_1), \dots, F_{dk}(x_d)\} + \sum_{j=1}^d \ln f_{jk}(x_j).$$

En supposant que les lois marginales sont les mêmes d'une classe à l'autre, on peut négliger le second terme et considérer simplement

$$g_k(\mathbf{x}) = \ln c_k \{F_1(x_1), \dots, F_d(x_d)\},$$

où F_1, \dots, F_d sont les marges *communes* aux K classes.

5.3 Modèle à base de copules Archimédiennes

5.3.1 Description générale

Rappelons qu'une copule d -dimensionnelle C est Archimédienne si elle s'écrit sous la forme

$$C_\varphi(u_1, \dots, u_d) = \varphi^{-1} \left\{ \sum_{j=1}^d \varphi(u_j) \right\},$$

où φ est un générateur. La densité d'une telle copule, si elle existe, est

$$c_\varphi(u_1, \dots, u_d) = (\varphi^{-1})^{(d)} \left\{ \sum_{j=1}^d \varphi(u_j) \right\} \prod_{j=1}^d \varphi'(u_j),$$

où φ' est la dérivée première de φ et $(\varphi^{-1})^{(p)}$ est la p -ème dérivée de φ^{-1} .

Dans le cas particulier où $d = 2$, on montre que

$$c_\varphi(u_1, u_2) = - \frac{\varphi'' \circ \varphi^{-1} \{ \varphi(u_1) + \varphi(u_2) \}}{[\varphi' \circ \varphi^{-1} \{ \varphi(u_1) + \varphi(u_2) \}]^3} \varphi'(u_1) \varphi'(u_2),$$

où φ'' est la dérivée seconde de φ . En supposant que les classes $\Omega_1, \dots, \Omega_K$ sont caractérisées par les copules Archimédiennes $C_{\varphi_1}, \dots, C_{\varphi_K}$, on a

$$g_k(\mathbf{x}) = (\varphi^{-1})^{(d)} \left[\sum_{j=1}^d \varphi \{ F_j(x_j) \} \right] \prod_{j=1}^d \varphi' \{ F_j(x_j) \}.$$

5.3.2 Exemple avec la copule de Clayton

Pour illustrer, considérons $K = 2$ classes Ω_1 et Ω_2 dont les structures de dépendance sont caractérisées respectivement par des copules de Clayton de paramètres θ_1 et θ_2 . La densité de cette copule pour $\theta \in [-1, \infty) \setminus \{0\}$ est

$$c(u_1, u_2) = (\theta + 1) (u_1 u_2)^{-(\theta+1)} (u_1^{-\theta} + u_2^{-\theta} - 1)^{-\left(\frac{1}{\theta}+2\right)}.$$

En supposant des marges uniformes sur $(0, 1)$, les fonctions discriminantes sont alors, en négligeant les termes constants,

$$g_1(x_1, x_2) = -(\theta_1 + 1) \ln(x_1 x_2) - \left(\frac{1}{\theta_1} + 2 \right) \ln(x_1^{-\theta_1} + x_2^{-\theta_1} - 1)$$

et

$$g_2(x_1, x_2) = -(\theta_2 + 1) \ln(x_1 x_2) - \left(\frac{1}{\theta_2} + 2 \right) \ln(x_1^{-\theta_2} + x_2^{-\theta_2} - 1).$$

Soit maintenant

$$h_{\theta_1, \theta_2}(x_1, x_2) = g_1(x_1, x_2) - g_2(x_1, x_2).$$

Une nouvelle observation $\mathbf{X} = (X_1, X_2)$ sera assignée à la classe Ω_1 dans le cas où $h_{\theta_1, \theta_2}(X_1, X_2) > 0$; au contraire, si $h_{\theta_1, \theta_2}(X_1, X_2) < 0$, elle est assignée à la classe Ω_2 . La courbe de x_2 en fonction de x_1 telle que $h_{\theta_1, \theta_2}(x_1, x_2) = 0$ partitionne le carré $[0, 1]^2$ en deux régions associées aux classes Ω_1 et Ω_2 . La Figure 5.1 présente cette courbe lorsque $\theta_1 = 2,9$ et $\theta_2 = 2,5$.

On remarque que la courbe est symétrique par rapport à la diagonale; ceci est mis en valeur par la droite d'équation $u = v$, présentée en bleu. Les valeurs à l'intérieur de la courbe en rouge correspondent à la classe Ω_1 . Le reste de la région $(0, 1)^2$ est composée des valeurs (x_1, x_2) qui mènent au choix de la classe Ω_2 . Ainsi, on peut dire que la classe Ω_2 serait sélectionnée lorsque les composantes d'un vecteur aléatoire ont tendance à être en *dépendance négative*, c'est-à-dire que les petites valeurs d'une des variables se présentent généralement accompagnées des grandes valeurs de l'autre. À l'inverse, la classe Ω_1 est caractérisée par une tendance à la *dépendance positive*. Ces conclusions sont cohérentes avec le fait que $\theta_1 = 2,9$ correspond à une copule

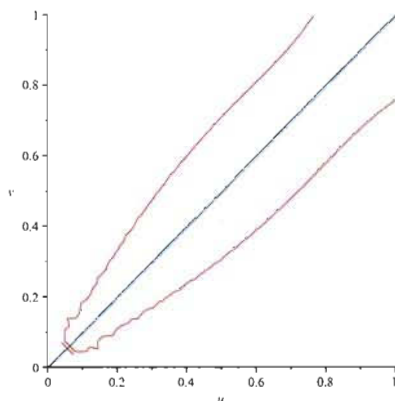


FIGURE 5.1: Seuil de décision pour deux copules de Clayton avec $\theta_1 = 2,9$ et $\theta_2 = 2,5$

dont la dépendance est plus forte que lorsque $\theta_2 = 2,5$. À noter finalement que la forme que l'on retrouve à la Figure 5.1 est typique de ce que l'on observe, peu importe les valeurs de θ_1 et de θ_2 .

5.3.3 Effet des marges sur l'affectation

On reprend l'exemple de la sous-section précédente, cette fois en posant $\theta_1 = \theta_2 = \theta$. Ici, plutôt que de supposer des marges uniformes sur $(0, 1)$, on suppose que F_1 et F_2 sont les fonctions de répartition de lois exponentielles de paramètres respectifs λ_1 et λ_2 . On a donc

$$F_1(x) = 1 - \exp\left(-\frac{x}{\lambda_1}\right) \quad \text{et} \quad F_2(x) = 1 - \exp\left(-\frac{x}{\lambda_2}\right).$$

On déduit alors

$$\begin{aligned} g_1(x_1, x_2) &= -(\theta + 1) \ln(F_1(x_1)F_1(x_2)) \\ &\quad - \left(\frac{1}{\theta} + 2\right) \ln(F_1(x_1)^{-\theta_1} + F_1(x_2)^{-\theta} - 1) \end{aligned}$$

et

$$\begin{aligned} g_2(x_1, x_2) &= -(\theta + 1) \ln(F_2(x_1)F_2(x_2)) \\ &\quad - \left(\frac{1}{\theta} + 2\right) \ln(F_2(x_1)^{-\theta_2} + F_2(x_2)^{-\theta} - 1.) \end{aligned}$$

Posons ensuite $h_{\lambda_1, \lambda_2}(x_1, x_2) = g_1(x_1, x_2) - g_2(x_1, x_2)$. Cette fois, la courbe des valeurs (x_1, x_2) telles que $h_{\lambda_1, \lambda_2}(x_1, x_2) = 0$ va partitionner la région $[0, \infty) \times [0, \infty)$ en deux sous-régions associées aux classes Ω_1 et Ω_2 . Une illustration est donnée à la Figure 5.2 lorsque $\lambda_1 = 1,8$, $\lambda_2 = 1,6$ et $\theta = 1,3$. Les valeurs (x_1, x_2) qui sont en-dessous de la courbe en rouge correspondent à la classe Ω_2 , alors que les valeurs au-dessus sont assignées à la classe Ω_1 . Ceci est cohérent avec le fait que la classe Ω_2 produit en moyenne des valeurs plus petites que la classe Ω_1 puisque $\lambda_2 < \lambda_1$.

5.4 Probabilités d'erreur de classification

Dans cette section, la probabilité d'un mauvais classement dans le contexte où les fonctions discriminantes sont représentées par des densités de copules sera examinée. On traitera d'abord du cas particulier où il y a deux classes ; une extension au cas à $K > 2$ classes sera ensuite présentée.

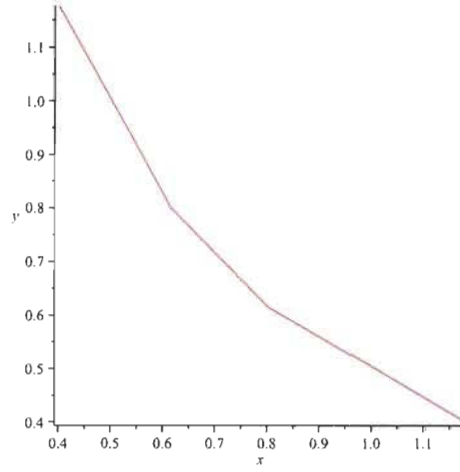


FIGURE 5.2: Seuil de décision pour deux copules de Clayton avec marges exponentielles

5.4.1 Cas à deux classes

Soient Ω_1 et Ω_2 , deux classes caractérisées par des régions \mathcal{R}_1 et \mathcal{R}_2 . Par définition, la probabilité d'une erreur de classement est donnée par

$$\begin{aligned}
 p_E &= P(\mathbf{x} \in \mathcal{R}_2 \cap \Omega_1) + P(\mathbf{x} \in \mathcal{R}_1 \cap \Omega_2) \\
 &= P(\mathbf{x} \in \mathcal{R}_2 | \Omega_1) P(\Omega_1) + P(\mathbf{x} \in \mathcal{R}_1 | \Omega_2) P(\Omega_2) \\
 &= \int_{\mathcal{R}_2} P(\mathbf{x} | \Omega_1) P(\Omega_1) d\mathbf{x} + \int_{\mathcal{R}_1} P(\mathbf{x} | \Omega_2) P(\Omega_2) d\mathbf{x}.
 \end{aligned}$$

Pour illustrer, reprenons l'exemple avec deux copules Archimédiennes de paramètres respectifs θ_1 et θ_2 . Si on suppose l'équiprobabilité des classes, c'est-

à-dire $P(\Omega_1) = P(\Omega_2) = 1/2$, alors on a

$$\begin{aligned} p_E &= \frac{1}{2} \int_{\mathcal{R}_2} (\theta_1 + 1) (x_1 x_2)^{-(\theta_1+1)} (x_1^{-\theta_1} + x_2^{-\theta_1} - 1)^{-\left(\frac{1}{\theta_1}+2\right)} dx_1 dx_2 \\ &\quad + \frac{1}{2} \int_{\mathcal{R}_1} (\theta_2 + 1) (x_1 x_2)^{-(\theta_2+1)} (x_1^{-\theta_2} + x_2^{-\theta_2} - 1)^{-\left(\frac{1}{\theta_2}+2\right)} dx_1 dx_2. \end{aligned}$$

À noter que les régions \mathcal{R}_1 et \mathcal{R}_2 sont délimitées par des fonctions f^- et f^+ qui sont symétriques par rapport à la droite d'équation $x_1 = x_2$, de sorte que

$$\mathcal{R}_1 = \{(x_1, x_2) \in (0, 1)^2 \mid f^-(x_1) \leq x_2 \leq f^+(x_1)\},$$

et $\mathcal{R}_2 = (0, 1)^2 \setminus \mathcal{R}_1$. Ainsi, on montre que

$$\begin{aligned} p_E &= \ell(\theta_1) \int_0^1 \left\{ \int_{f^-}^{f^+} \{x\phi(x)\}^{-(\theta_1+1)} \{q(x, \theta_1)\}^{-\left(\frac{1}{\theta_1}+2\right)} dx_2 \right\} dx_1 \\ &\quad + \ell(\theta_2) \int_0^1 \left\{ \int_0^{f^-} \{x\phi(x)\}^{-(\theta_2+1)} \{q(x, \theta_2)\}^{-\left(\frac{1}{\theta_2}+2\right)} dx_2 \right\} dx_1 \\ &\quad + \ell(\theta_2) \int_0^1 \left\{ \int_{f^+}^1 \{x\phi(x)\}^{-(\theta_2+1)} \{q(x, \theta_2)\}^{-\left(\frac{1}{\theta_2}+2\right)} dx_2 \right\} dx_1, \end{aligned}$$

où $\ell(\theta) = (\theta+1)/2$ et $q(x, \theta) = x^{-\theta} + \{\phi(x)\}^{-\theta} - 1$. Comme les fonctions f^- et f^+ sont implicites, il n'est pas possible d'obtenir une expression explicite pour cette probabilité d'erreur de classification. Il est cependant possible d'approcher cette valeur par simulations. À cette fin, soient $(X_1, Y_1), \dots, (X_n, Y_n)$, des copies aléatoires de (X, Y) de loi C . D'après la loi forte des grands nombres,

$$\frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[(X_i, Y_i) \in \mathcal{A}]} \xrightarrow{n \rightarrow \infty} \mathbb{E} \{ \mathbf{1}_{[(X, Y) \in \mathcal{A}]} \} = P((X, Y) \in \mathcal{A}).$$

On peut ainsi estimer la probabilité d'erreur de classification en écrivant

$$p_E = P(\mathbf{X} \in \mathcal{R}_1 | \Omega_2) P(\Omega_2) + P(\mathbf{X} \in \mathcal{R}_2 | \Omega_1) P(\Omega_1). \quad (5.2)$$

Soit donc l'algorithme d'approximation suivant afin d'estimer p_E .

1. Pour $j = 1, 2$, simuler un échantillon de taille n à partir de la copule C_j ; on note cet échantillon

$$\left(X_1^{(j)}, Y_1^{(j)}\right), \dots, \left(X_n^{(j)}, Y_n^{(j)}\right).$$

2. Afin de pallier au problème des marges inconnues, transformer les échantillons générés en rangs $(R_i^{(1)}, S_i^{(1)})$ et $(R_i^{(2)}, S_i^{(2)})$, $i \in \{1, \dots, n\}$.
3. Les probabilités $P(\mathcal{R}_2|\Omega_1)$ et $P(\mathcal{R}_1|\Omega_2)$ sont alors estimées par

$$P(\widehat{\mathcal{R}_2|\Omega_1}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[h(R_i^{(1)}, S_i^{(1)}) < 0]}$$

et

$$P(\widehat{\mathcal{R}_1|\Omega_2}) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[h(R_i^{(2)}, S_i^{(2)}) > 0]},$$

où

$$g(R_i^{(1)}, S_i^{(1)}) = c_1(R_i^{(1)}, S_i^{(1)}, \theta_1) - c_2(R_i^{(1)}, S_i^{(1)}, \theta_2)$$

et

$$g(R_i^{(2)}, S_i^{(2)}) = c_1(R_i^{(2)}, S_i^{(2)}, \theta_1) - c_2(R_i^{(2)}, S_i^{(2)}, \theta_2).$$

4. Remplacer ensuite les probabilités calculées à l'étape précédente dans la formule (5.2) en prenant $P(\Omega_1) = P(\Omega_2) = 1/2$.

En générant $n = 1\,000$ couples (X, Y) à partir de copules de Clayton avec des tau de Kendall différents, on obtient les résultats présentés au Tableau 5.1. Ces résultats montrent bien que plus les classes ont un tau de Kendall semblable, plus l'erreur d'une mauvaise classification est grande. Au contraire, plus les classes sont différentes, en d'autres termes plus elles ont des tau de Kendall différents, plus l'erreur est faible.

TABLE 5.1: Estimation des probabilités d'erreurs de classification pour la copule de Clayton

Valeur de τ_1	Valeur de τ_2			
	0.25	0.50	0.75	1
0.25	—	0.405	0.266	0.005
0.50	0.395	—	0.329	0.007
0.75	0.266	0.358	—	0.008
1	0.005	0.006	0.008	—

5.4.2 Cas à K classes

Dans le cas où nous avons plus de deux classes, il est aussi possible de trouver cette probabilité d'erreur de classement. Cependant, la probabilité d'erreur est plus importante et plus difficile à calculer du fait du nombre élevé de classes. Ainsi, il est plus simple de considérer la probabilité d'un bon classement, à savoir (?)

$$\begin{aligned}
 1 - p_E &= \sum_{k=1}^K P(\mathbf{X} \in \mathcal{R}_k, \Omega_k) \\
 &= \sum_{k=1}^K P(\mathbf{X} \in \mathcal{R}_k | \Omega_k) P(\Omega_k) \\
 &= \sum_{k=1}^K \int_{\mathcal{R}_k} p(\mathbf{x} | \Omega_k) P(\Omega_k) d\mathbf{x}.
 \end{aligned}$$

En partant du même principe que pour la probabilité d'erreur appliquée au cas bivarié, le but de définir chacune des régions \mathcal{R}_k via

$$\mathcal{R}_k := \{\mathbf{x} | h_k(\mathbf{x}) > 0\},$$

où

$$h_k(\mathbf{x}) = P(\mathbf{x}|\Omega_k) - \max_{k' \neq k} P(\mathbf{x}|\Omega_{k'}).$$

Finalement, la probabilité d'un bon classement est

$$1 - p_E = \sum_{k=1}^K \int_{\mathcal{R}_k} p(\mathbf{x}|\Omega_k) P(\Omega_k) d\mathbf{x}.$$

Cette probabilité peut être estimée par simulation, à l'instar du cas bivarié.

5.5 Illustration sur de vraies données

Le jeu de données utilisé pour appliquer notre méthode sur des données réelles est extrait d'une étude réalisée sur des images de billets de banque par Volker Lohweg et Helene Doerksen, tous deux chercheurs à l'Université des Sciences Appliquées d'Ostwestfalen-Lippe en Allemagne. Les données ont été téléchargées sur le site Web de K. Bache et M. Lichman. Ce fichier contient deux variables quantitatives caractérisant les images. La première correspond à l'entropie de l'image, qui mesure la complexité de l'image en terme de nombre de pixels différents. La seconde est la variance de la transformation en ondelettes de l'image, qui permet de connaître l'intensité des variations des oscillations de l'image. Une troisième variable qualitative permettant de déterminer si un billet de banque est falsifié complète le jeu de données.

Le jeu de données contient 1 372 observations ; 45% appartiennent à la Classe 1 des billets falsifiés, alors que 55% appartiennent à la Classe 2 des billets authentiques. Afin de vérifier l'efficacité de notre méthode, on a scindé le jeu de données en deux. Ainsi, on utilisera les 1 000 premières observations

pour valider le modèle sous-jacent, alors que les 372 observations restantes seront classées selon notre méthode de classification à base de copules.

La première étape avant de lancer l'algorithme est de calculer les tau de Kendall correspondant aux deux classes. Ainsi, lorsque la variable de groupe vaut 0, alors le tau de Kendall empirique vaut $\tau_{n,0} = 0,2659$; lorsque la variable de groupe vaut 1, on a $\tau_{n,1} = 0,1830$.

La seconde étape consiste à valider le choix de la copule qui sera utilisée pour la classification. En appliquant le test d'adéquation développé par ?, on conclut que la copule de Clayton est un modèle raisonnable pour les échantillons des deux classes. Également, à la lumière des nuages de points des rangs standardisés présentés à la Figure 5.3, on aurait pu arriver à cette conclusion. En effet, on remarque une forte concentration de paires de points dans le carré inférieur gauche, ce qui est typique de la densité de la copule de Clayton présentée au Chapitre 4.

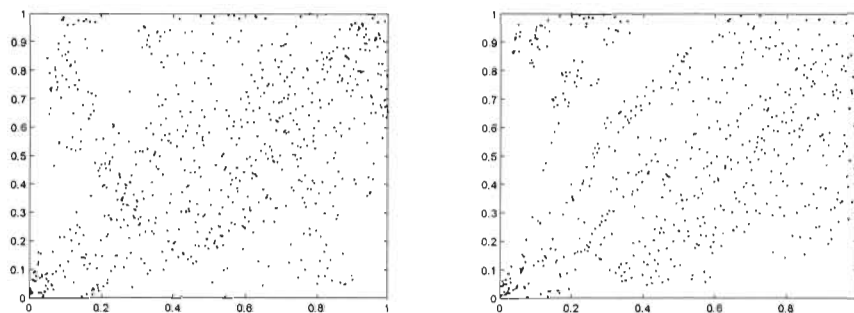


FIGURE 5.3: Nuage de points des rangs standardisés du groupe Ω_1 (à gauche) et Ω_2 (à droite)

Une fois ces paramètres déterminés, nous exécutons l'algorithme de classification. Il s'agit donc, pour une nouvelle observation $\tilde{\mathbf{X}} = (\tilde{X}_1, \tilde{X}_2)$, de déterminer sa classe d'appartenance parmi Ω_1 et Ω_2 . Pour effectuer cette classification, on transforme d'abord $\tilde{\mathbf{X}}$ en rangs. Ainsi, la fonction discriminante sera calculée à partir de $\tilde{\mathbf{R}} = (\tilde{R}_1, \tilde{R}_2)$, où \tilde{R}_1 est la position (le rang) de \tilde{X}_1 parmi les valeurs de la première variable de l'échantillon de validation ; de même, \tilde{R}_2 est la position (le rang) de \tilde{X}_2 parmi les valeurs de la deuxième variable de l'échantillon de validation. La nouvelle observation $\tilde{\mathbf{X}}$ est affectée à la classe Ω_1 si la valeur de la fonction discriminante associée est supérieure à la valeur de celle de la classe Ω_2 ; elle est affectée à Ω_2 dans le cas contraire.

Puisque l'on connaît les vraies classes d'appartenance des 372 observations restantes, on peut évaluer la capacité de notre méthode à effectuer une bonne classification. Ainsi, la méthode de classification a attribué à la bonne classe 237 de ces *nouvelles observations*, pour un pourcentage de réussite de $100 \times 237/372 = 63\%$. Ce taux de succès peut paraître peu élevé. Cependant, compte tenue de la faible différence entre les tau de Kendall des deux classes, on arrive néanmoins à la conclusion que notre méthode donne des résultats intéressants et cohérents.

CHAPITRE 6

CONCLUSION

Les méthodes de classification telles que la classification Bayésienne sont très souvent utilisées. Cependant, il est fréquent d'être confronté à de graves erreurs d'attribution de classes lorsque l'hypothèse de normalité n'est pas respectée. Ce mémoire avait pour but de construire une nouvelle méthode de classification en combinant l'approche avec la théorie des copules. Cette théorie est principalement basée sur le fait qu'il est possible de modéliser une distribution sans tenir compte des comportements marginaux, à partir de la copule qui lui est associée. Cette copule remplace alors la densité de la loi Normale utilisée dans la méthode de classification Bayésienne.

La première démarche de ce mémoire était de prouver mathématiquement, à partir de la théorie des copules, que cette méthode est applicable. Une fois notre théorie démontrée, nous avons procédé à des simulations de classification de données générées à partir de la densité de la copule de Clayton. Le but étant de calculer les taux d'erreur qu'engendre une telle méthode pour plusieurs valeurs de tau de Kendall. Cette étape nous a permis d'appuyer la démonstration mathématique et d'aborder la dernière phase de cette recherche, à savoir l'application sur de vraies données. En effet, afin d'étayer

notre théorie, nous nous sommes appuyés sur un exemple concret de classification basé sur un jeu de données mis à la disposition gracieusement par des chercheurs d'une université allemande. L'illustration montre assez bien l'efficacité de la méthode introduite, bien que la différence entre les deux classes était relativement faible.

Dans le futur, il serait intéressant de pousser plus loin l'idée proposée dans ce mémoire. En effet, plusieurs points ont été évoqués rapidement et méritent d'être approfondis. Premièrement, pour développer cette nouvelle méthode de classification, nous nous sommes uniquement penchés sur le cas de la copule de Clayton. Or, la famille des copules Archimédiennes est composée de bien d'autres copules dont il faudrait étudier le comportement vis-à-vis cette nouvelle approche. De même, bien que la famille des copules Archimédiennes est l'une des plus citées dans la littérature, d'autres familles pourraient être étudiées, par exemple la famille des copules elliptiques. Une approche complémentaire pourrait alors être abordée dans l'optique d'adapter la présente méthode à l'ensemble des familles de copules existantes. Deuxièmement, nous nous sommes limités à observer le comportement de cette approche pour le cas bidimensionnel à $K = 2$ classes. Un prochain travail pourrait donc être élaboré pour décrire plus en détail le comportement des copules appliquées à la méthode de classification avec d variables. Pour finir, il serait intéressant de pouvoir faire des comparaisons entre ce nouveau modèle et les modèles classiques de classifications supervisées pour déterminer à quel point l'approche présentée ici est performante.

Annexe A

PROGRAMMES DE SIMULATION MATLAB

A.1 Estimation des probabilités d'erreur

```
1 function erreur = Proba_erreur(n,N,Copule,tau1,tau2)
2
3 X1 = sim_copula(n,Copule,tau1);
4 X2 = sim_copula(n,Copule,tau2);
5 R1 = tiedrank(X1(:,1)) / n;
6 S1 = tiedrank(X1(:,2)) / n;
7 R2 = tiedrank(X2(:,1)) / n;
8 S2 = tiedrank(X2(:,2)) / n;
9
10 somme=0; erreur_serie=0;
11
12 theta1 = tau_theta(1,tau1); theta2 = tau_theta(1,tau2);
13 i=1; j=1;
14 for i=1:N
15     c1=0;
16     c2=0;
17     for j=1:n
18         v = fonct_discrim(R1(j),S1(j),theta1,theta2);
19         if v < 0
20             c1 = c1+1;
21         end
22
23         v = fonct_discrim(R2(j),S2(j),theta1,theta2);
24         if v > 0
25             c2 = c2+1;
26         end
27     end
28
29     erreur_serie = (erreur_serie * 0)+(c1/n)*0.5 + (c2/n)*0.5;
30     somme = somme + erreur_serie;
31 end
32
33 erreur = somme/N;
```

A.2 Estimation de paramètres par inversion du tau de Kendall

```

1 function theta = tau_theta(model,tau)
2
3 %Returns the parameter of a copula which corresponds to
4 %a given value of Kendall's tau
5 %1. Clayton
6 %3. Ali-Mikhail-Haq
7 %4. Gumbel
8 %5. Frank
9 %7. Farlie-Gumbel-Morgenstern
10 %8. Plackett
11 %9. Gumbel A
12 %11. Normal
13 %12. Student
14 %13. Frechet
15 %15. Galambos
16 %16. Cuadras-Auge
17 %17. Ordinal sum
18 %18. Durante A
19 %19. Durante B
20
21 %Clayton
22 if (model == 1)
23     if tau == 1
24         theta = 1000;
25     else theta = 2*tau/(1-tau);
26     end
27
28 %Ali-Mikhail-Haq
29 elseif (model == 3)
30     if (tau <= 0.0)
31         theta = 0;
32     elseif (tau > .33)
33         theta = 1;
34     else
35         f = @(x)((3*x^2-2*x-2*(1-x)^2*log(1-x))/(3*x^2)) - tau);
36         theta = fzero(f,.1);
37     end
38
39 %Gumbel
40 elseif (model == 4)
41     if (tau <= 0)
42         theta = 0;
43     else
44         theta = tau;
45     end
46
47 %Frank
48 elseif (model == 5)

```

```

49     fp1 = fopen('tau_Frank.txt', 'r');
50     b2 = -1; t2 = 0;
51     while (tau > b2)
52         t1 = t2;
53         b1 = b2;
54         a = fscanf(fp1, '%f %f', 2);
55         b2 = a(1); t2 = a(2);
56     end
57     fclose(fp1);
58     theta = ((t1 - t2)*tau + (b1*t2 - b2*t1))/(b1 - b2);
59
60 %Farlie-Gumbel-Morgenstern
61 elseif (model == 7)
62     if (tau < -2/9)
63         theta = -1;
64     elseif (tau > 2/9)
65         theta = 1;
66     else
67         theta = 9*tau/2;
68     end
69
70 %Plackett
71 elseif (model == 8)
72     if (tau <= 0)
73         theta = 0;
74     else
75         fp1 = fopen('tau_Plackett.txt', 'r');
76         b2 = -1;
77         t2 = 0;
78         while (tau > b2)
79             t1 = t2;
80             b1 = b2;
81             a = fscanf(fp1, '%f %f', 2);
82             b2 = a(1); t2 = a(2);
83         end
84         fclose(fp1);
85         theta = ((t1 - t2)*tau + (b1*t2 - b2*t1))/(b1 - b2);
86     end
87
88 %Gumbel A
89 elseif (model == 9)
90     if (tau <= 0)
91         theta = 0;
92     elseif (tau > 0.4184)
93         theta = 1;
94     elseif ( (tau <= 0.4184) && (tau > 0) )
95         f = @(x)((8*atan(sqrt(x/(4-x)))/sqrt(x*(4-x))) - 2 - tau);
96         theta = fzero(f,.5);
97     end
98
99 %Marshall-Olkin (MO)
100 elseif (model == 10)

```

```

101     if (tau <= 0)
102         theta = 0;
103     else
104         theta = 2*tau/(1+tau);
105     end
106
107 % Normal
108 elseif (model == 11)
109     theta = sin((pi/2)*tau);
110
111 % Frechet
112 elseif (model == 13)
113     if (tau < 0)
114         theta = 0;
115     else
116         theta = realsqrt((3*tau)+1) - 1;
117     end
118
119 % Galambos
120 elseif (model == 15)
121     if (tau <= 0)
122         theta = 0;
123     else
124         fp1 = fopen('tau_Galambos.txt', 'r');
125         b2 = -1; t2 = 0;
126         while (tau > t2)
127             t1 = t2;
128             b1 = b2;
129             a = fscanf(fp1, '%f %f', 2);
130             b2 = a(1); t2 = a(2);
131         end
132         fclose(fp1);
133         theta = ((b1 - b2)*tau + (t1*b2 - t2*b1))/(t1 - t2);
134     end
135
136 % Cuadras-Auge
137 elseif (model == 16)
138     if (tau < 0)
139         theta = 0;
140     else
141         theta = 2*tau/(1+tau);
142     end
143
144 % Ordinal sum
145 elseif (model == 17)
146     theta = 1/sqrt(1-max(0,tau));
147
148 % Durante A
149 elseif (model == 18)
150     if (tau <= 0)
151         theta = 1;
152     else

```

```

153     f = @(x)(1 - (2*(7*x+2))/((x+1)*(x+2)*(2*x+1)) - tau);
154     if (tau < .85)
155         theta = fzero(f,2);
156     elseif (tau >= .85 && tau <=.94)
157         theta = fzero(f,5);
158     elseif (tau >.94)
159         theta = fzero(f,12);
160     end
161 end
162
163 %Durante B
164 elseif (model == 19)
165     if (tau <= 0)
166         theta = 0;
167     else
168         s = 0.00001; v = 0.9; v1 = 0.1;
169         while ( abs(v-v1) > s )
170             t2 = (v^4 - 2*v^3 - 9*v^2 - 6*v + 6*(v^2+2*v+1)*log(v+1))/(2*v^4.) - tau;
171             t1 = (v1^4 - 2*v1^3 - 9*v1^2 - 6*v1 + 6*(v1^2+2*v1+1)*log(v1+1))/(2*v1^4) - tau;
172             v2 = v1 - t1*(v1-v)/(t1-t2);
173             if (v2 <= 0)
174                 v2 = 0.001;
175             elseif (v2 >= 1)
176                 v2 = 0.999;
177             end
178             v = v1;
179             v1 = v2;
180             if ( abs(v1) > 100000 )
181                 v = tau;
182                 v1 = tau-0.01;
183             end
184         end
185         theta = v1;
186     end
187
188 %Student
189 elseif (model <= 40 && model>= 21)
190     theta = sin((pi/2)*tau);
191
192 end

```


A.3 Simulation de données bivariées selon une certaine copule

```

1 function X = sim_copula(n,model,tau)
2
3 % Generates a bivariate sample of size n from the following copulas :
4 % 1. Clayton
5 % 3. Ali-Mikhail-Haq
6 % 4. Gumbel-Hougaard (extreme)
7 % 5. Frank
8 % 7. Farlie-Gumbel-Morgenstern
9 % 8. Plackett
10 % 9. Gumbel A (extreme)
11 % 10. Marshall-Olkin (extreme)
12 % 11. Normal
13 % 12. Student
14 % 13. Fréchet
15 % 16. Cuadras-Augé (extreme)
16 % 17. Ordinal sum
17 % 18. Durante A
18 % 19. Durante B
19 % 21-40. Student with "model-20" degrees of freedom
20 %
21 % author : Jean-François Quessy, Ph. D.
22
23 X = zeros(n,2);
24
25 theta = tau_theta(model,tau);
26
27 % Clayton
28 if (model == 1)
29     if (theta == 0)
30         X = rand(n,2);
31     else
32         for i=1:n
33             v = rand;
34             t = rand;
35             a = v^(-theta);
36             b = t^(-theta/(theta+1)) - 1;
37             X(i,1) = (a*b+1)^(-1/theta);
38             X(i,2) = v;
39         end
40     end
41
42 % Ali-Mikhail-Haq
43 elseif (model == 3)
44     if (theta == 0)
45         X = rand(n,2);
46     else
47         for i=1:n
48             u = rand; t = rand;

```

```

49     a = -theta - 4*t*theta^2*u + 2*theta*t*u - 2*theta*t
50         + 2*theta^2*t + 2*theta^2*t*u^2 + 1;
51     b = 1 - 2*theta + 4*t*theta^2*u^2 + 4*theta*t*u
52         - 4*theta^2*t*u + theta^2;
53     c = -2*theta^2*t*u + theta^2*u^2*t - theta + theta^2*t;
54     X(i,1) = u;
55     X(i,2) = (a-sqrt(b))/(2*c);
56   end
57 end
58
59 % Gumbel-Hougaard
60 elseif (model == 4)
61     if (theta == 0)
62         X = rand(n,2);
63     else
64         for i=1:n
65             U0 = rand; U1 = rand; U2 = rand;
66             z = (U0^(1-theta))/(U0^(1-theta)+(1.0-U0)^(1-theta));
67             if (rand < 1-theta)
68                 w = U1*U2;
69             else
70                 w = U2;
71             end
72             t1 = ((1-z)^(1/(1-theta)) + z^(1/(1-theta)))^(1-theta);
73             X(i,1) = w^(z/t1);
74             X(i,2) = w^((1.0-z)/t1);
75         end
76     end
77
78 % Frank
79 elseif (model == 5)
80     if (theta == 0)
81         X = rand(n,2);
82     else
83         for i=1:n
84             u = rand; t = rand;
85             t1 = t*exp(-theta) + (1-t)*exp(-theta*u);
86             t2 = t + (1-t)*exp(-theta*u);
87             X(i,1) = u;
88             X(i,2) = log(t2/t1)/theta;
89         end
90     end
91
92 % Farlie-Gumbel-Morgenstern
93 elseif (model == 7)
94     if (theta == 0)
95         X = rand(n,2);
96     else
97         for i=1:n
98             u = rand; t = rand;
99             b = theta*(1-2*u);
100             root = sqrt((1+b)^2 - 4*t*b);

```

```

101         X(i,1) = u;
102         X(i,2) = 2*t/(1+b+root);
103     end
104 end
105
106 %Plackett
107 elseif (model == 8)
108     if (theta == 0)
109         X = rand(n,2);
110     else
111         for i=1:n
112             u = rand; t = rand;
113             a = t*(1-t);
114             b = theta + a*(theta-1)^2;
115             c = 2*a*(u*theta^2 + 1 - u) + theta*(1-2*a);
116             d = sqrt(theta)*sqrt(theta + 4*a*u*(1-u)*(1-theta)^2);
117             X(i,1) = u;
118             X(i,2) = (c-(1-2*t)*d)/(2*b);
119         end
120     end
121
122 %Gumbel A
123 elseif (model == 9)
124     c = (4+theta)/(4-theta);
125     for i=1:n
126         U1 = unifrnd(0,1); U2 = unifrnd(0,1); U3 = unifrnd(0,1);
127         %rejection method
128         Z = 0; Z1 = 1; % arbitrarily
129         while (Z ~= Z1)
130             Z1 = unifrnd(0,1); U0 = unifrnd(0,1);
131             t1 = ( 2 - theta*(2*Z1^2 - 2*Z1 + 1) )
132                 / ( (theta*Z1^2 - theta*Z1 + 1)^2 ) - 1;
133             if (U0<=t1/c)
134                 Z = Z1;
135             end
136         end
137
138         t2 = 2*theta*Z*(1-Z)*(theta*Z^2 - theta*Z + 1);
139         t3 = 2 - 2*theta*Z^2 + 2*theta*Z - theta
140             - (theta*Z^2 - theta*Z + 1)^2;
141         t4 = t2/t3;
142         if (U3<=t4)
143             W = U1;
144         else
145             W = U1*U2;
146         end
147         t5 = theta*Z^2 - theta*Z + 1;
148         X(i,1) = W^(Z/t5);
149         X(i,2) = W^((1-Z)/t5);
150     end
151
152 % Marshall-Olkin (equal parameters)

```

```

153 elseif (model == 10)
154     if (theta == 0)
155         X = rand(n,2);
156     else
157         for i=1:n
158             U = unifrnd(0,1,1,3);
159             X(i,1) = max(U(1,1)^(1/(1-theta)),U(1,3)^(1/theta));
160             X(i,2) = max(U(1,2)^(1/(1-theta)),U(1,3)^(1/theta));
161         end
162     end
163
164 %Normal
165 elseif (model == 11)
166     mu = [0 0]; sigma = [1 theta ; theta 1];
167     for i=1:n
168         X(i,:) = mvnrnd(mu,sigma);
169     end
170     X = normcdf(X);
171
172 %Fréchet
173 elseif (model == 13)
174     if (theta == 0.0)
175         X = rand(n,2);
176     else
177         for i=1:n
178             t = rand; s = rand; X(i,2) = s;
179             if (t <= theta)
180                 X(i,1) = s;
181             else
182                 X(i,1) = rand;
183             end
184         end
185     end
186
187 %Cuadras-Augé
188 elseif (model == 16)
189     if (theta == 0)
190         X = rand(n,2);
191     else
192         for i=1:n
193             t = rand; v = rand; X(i,2) = v;
194             if ( t < (1-theta)*v^(1-theta) )
195                 X(i,1) = t*v^theta/(1-theta);
196             elseif (t < v^(1-theta))
197                 X(i,1) = v;
198             else
199                 X(i,1) = t^(1/(1-theta));
200             end
201         end
202     end
203
204 %Ordinal sum

```

```

205 elseif (model == 17)
206     for i=1:n
207         t = rand; v = rand; X(i,2) = v;
208         if ( v <= 1/theta )
209             X(i,1) = t/theta;
210         else
211             X(i,1) = v;
212         end
213     end
214
215 %Durante A
216 elseif (model == 18)
217     if (theta == 1.0)
218         X = rand(n,2);
219     else
220         for i=1:n
221             t = rand; v = rand; X(i,2) = v;
222             if ( t < theta*v*(1-v)^(theta-1) )
223                 X(i,1) = t/(theta*(1.0-v)^(theta-1.0));
224             elseif ( t < 1.0 - (1.0-v)^theta )
225                 X(i,1) = v;
226             else
227                 X(i,1) = 1.0 - (1.0-t)^(1.0/theta);
228             end
229         end
230     end
231
232 %Durante B
233 elseif (model == 19)
234     if (theta == 0.0)
235         X = rand(n,2);
236     else
237         for i=1:n
238             t = rand; v = rand; X(i,2) = v;
239             if ( t < (theta+1)*v/(theta*v+1)^2 )
240                 X(i,1) = (theta*v+1)^2*t/(theta+1);
241             elseif ( t < (theta+1)*v/(theta*v+1) )
242                 X(i,1) = v;
243             else
244                 X(i,1) = t/(1+theta-theta*t);
245             end
246         end
247     end
248
249 %Student
250 elseif (model <= 40 && model >= 21)
251     sigma = [1 theta ; theta 1];
252     for i=1:n
253         X(i,:) = mvtrnd(sigma,model-20);
254     end
255     X = tcdf(X,model-20);
256 end

```

A.4 Transformation d'un échantillon univarié en rangs

```
1 function R = rankit(X)
2
3 R = tiedrank(X);
```

A.5 Calcul des fonctions discriminantes

```

1 function p = fonct_discrim(u,v,theta1,theta2)
2
3 % fonction permettant de calculer la différence entre
4 % deux densités de copules de Clayton
5
6 p = ((theta1+1)*(u*v)^(-theta1-1)*( u^(-theta1) + v^(-theta1) - 1)^((-1/theta1)-2))
7      - ((theta2 + 1)*(u*v)^(-theta2-1)*( u^(-theta2) + v^(-theta2) - 1)^((-1/theta2)-2));

```

A.6 Classification d'une nouvelle observation

```

1 function Classe = Classif(X,Y)
2
3 % X : jeu de données de référence
4 % Y : une nouvelle paire d'observations
5
6 % Séparation de la matrice en deux, en fonction des classes
7 X0 = X(X(:,3)==0,1:2); X1 = X(X(:,3)==1,1:2);
8
9 % Détermination du rang de Y(1) parmi X0(:,1) et du rang de Y(2) parmi X0(:,2)
10 R0 = [ sum(X0(:,1)<=Y(1)) sum(X0(:,2)<=Y(2)) ] / length(X0);
11 % Détermination du rang de Y(1) parmi X1(:,1) et du rang de Y(2) parmi X1(:,2)
12 R1 = [ sum(X1(:,1)<=Y(1)) sum(X1(:,2)<=Y(2)) ] / length(X1);
13
14 % Calcul du tau de Kendall et estimation de theta
15 tau0 = tau(X0);
16 %tau0=0.2659;
17 theta0 = tau_theta(1,tau0);
18 tau1 = tau(X1);
19 %tau1=0.1830;
20 theta1 = tau_theta(1,tau1);
21
22 % Calcul de g1 et de g2
23 g0 = densite_clayton(R0,theta0);
24 g1 = densite_clayton(R1,theta1);
25
26 % Détermination de la classe d'appartenance
27 if (g0 > g1)
28     Classe = 0;
29 else
30     Classe = 1;
31 end

```

A.7 Calcul de la fonction discriminante

```
1 function p = densite_clayton(x,theta)
2
3 %Fonction permettant de calculer la densité de la copule de Clayton
4 u = x(1); v = x(2);
5 p = ((theta+1)*(u*v)^(-theta-1)*( u^(-theta) + v^(-theta) - 1)^((-1/theta)-2)) ;
```


A.8 Algorithme de classification d'un jeu de données

```
1 function resultat = ParcourNouvelleObs(X,Z)
2 % Fonction permettant de classer l'ensemble des nouvelles observations et
3 % de les regrouper dans un vecteur. Elle renvoie une matrice Nx4, où la classe
4 % attribuée est en quatrième position.
5 % X est l'échantillon témoin et Z l'échantillon test.
6 N=length(Z);
7 Y = (1:2);
8 Res=(1:N);
9 R=Res';
10
11 for i=1:N
12     Y(1)=Z(i,1);
13     Y(2)=Z(i,2);
14     R(i)=Classif(X,Y);
15 end
16
17 resultat=[Z,R];
```

A.9 Calcul du taux d'erreur de classification

```
1 function erreur = ErreurClassification(X)
2
3 E = 0;
4 N = length(X);
5
6 for j=1:N
7     if X(j,3)~= X(j,4)
8         E=E+1;
9     end
10    j=j+1;
11 end
12
13 erreur = E/N*100;
```